DOCUMENT RESUME

ED 031 124 24 FL 001 391

By-Politzer, Robert L.; Weiss, Louis
Characteristics and Behaviors of the Successful Foreign Language Teacher. Technical Report No. 5.
Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
Spons Agency-Office of Education (DHEW), Washington, D.C.
Report No-SCRDT-TR-5
Bureau No-BR-5-0252
Pub Date Apr 69
Contract-OEC-6-10-078
Note-114p.
EDRS Price MF-$0.50 HC-$5.80
Descriptors-*Academic Achievement, *Behavioral Science Research, Classroom Observation Techniques, Data Analysis, Effective Teaching, French, Language Instruction, *Language Teachers, Language Tests, Questionnaires, *Secondary School Students, Secondary School Teachers, Student Attitudes, Student Motivation, Teacher Behavior, Teacher Characteristics, Teacher Evaluation, *Teacher Influence, Video Tape Recordings
Identifiers-California, San Jose

An experiment with 17 teachers of first-year high school French recruited from the staffs of three large, diversified San Jose, California school districts was designed to (1) observe closely the classroom behaviors of the participating teachers, (2) identify those successful in terms of pupil achievement, and (3) compare the behaviors and characteristics of these teachers with those identified as less successful. The predominant feature of the research was the systematic observation and rating of the classroom behavior of the teachers by recording four 15-minute class segments over a period of five months. By means of specially constructed criterion tests, student and teacher questionnaires, the Modern Language Association Proficiency Tests for Teachers, and the observation of classroom procedures, data were gathered that yielded 40 variables. An analysis of the intercorrelation matrix of the 40 variables showed that a series of interrelated teacher behaviors and characteristics correlated significantly with student achievement. The two research hypotheses formulated focused on the linguistically definable and pedagogical aspects of the teachers's classroom behavior. The major recommendations emerging from the study concerned teacher training innovations and further research needs. (Author/AF)

ED031124

BR-5-0252
PA 24
OE-BR

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Technical Report No. 5

CHARACTERISTICS AND BEHAVIORS OF THE SUCCESSFUL FOREIGN LANGUAGE TEACHER

Robert L. Politzer and Louis Weiss

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

School of Education
Stanford University
Stanford, California

April 1969

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office of Education. (Contract No. OEC-6-10-078, Project 5-0252-0103.)

FL001391

Acknowledgments

The authors wish to express their thanks and acknowledge their indebtedness to all those whose help and cooperation made this study possible. Particular thanks are due to:

The French teachers and their students who were the subjects of this study.

The administrators who allowed the use of videotape in the classrooms under their jurisdiction.

The teachers who helped in the rating and observation of classroom behaviors and the grading of the criterion tests.

Dr. G. Richard Tucker of McGill University for reading the entire manuscript and making many valuable suggestions.

The personnel of the Methodology Unit of the Stanford Center for Research and Development in Teaching, in particular Dr. Janet D. Elashoff, for their invaluable help with the statistical interpretation of the data.

Robert L. Politzer
Louis Weiss

Abstract

Improvements in foreign language education have generally been concerned with the production of new teaching materials and the upgrading of teacher training programs. Comparatively little research has studied the relationship between the teacher's knowledge and skills and student achievement. The main purpose of this study was to observe as closely as possible the classroom behaviors of a group of teachers, to identify those who were successful in terms of pupil achievement, and to compare the behaviors and characteristics of these teachers with those of the teachers who were identified as less successful.

The classroom behaviors of 17 teachers of first-year high school French were recorded on videotape in four 15-minute segments over a period of five months. By means of specially constructed criterion tests, student and teacher questionnaires, MLA-ETS Language Proficiency Tests for Teachers, and systematic observation and rating of classroom procedures, data were gathered that yielded 40 variables. Class means on the criterion tests were adjusted for aptitude as measured by the Carroll-Sapon MLAT, administered at the beginning of the study. In addition to an analysis of the correlation matrices, two research hypotheses were formulated, one focusing on a linguistically definable aspect of the teacher's classroom behavior and the other on a pedagogical aspect.

The teachers participating in the study were recruited from the staffs of three large school districts in the San Jose, California area. The general population in this area covers a broad socio-economic spectrum from auto and aircraft factory workers to professional workers in electronics and communications systems firms. A survey of the student population in the participating schools showed a range of college aspiration from 93% to 30%.

The first hypothesis, which predicted a significantly higher level of achievement in the five classes whose teachers ranked highest in the proportion of free to controlled drill activities than in the five classes whose teachers ranked lowest on this variable, was not sustained. The second hypothesis, which predicted that teachers who vary classroom procedures from free to controlled drills would be more successful than those who stay with the same type of drill for prolonged periods of time, was sustained.

An analysis of the intercorrelation matrix of the 40 variables used in the study showed that a series of interrelated teacher behaviors and characteristics correlated significantly with student achievement. Among these characteristics were the use of free response drill, the use of visual aids, frequency of switching from one type of drill to another, residence in France, and high performance in aural comprehension. The conclusion is thus advanced that residence in the foreign country and a certain propensity for innovation and flexibility represent desirable characteristics of the foreign language teacher.

The major recommendations emerging from this study concern teacher training as well as further research. Teacher training should be concerned with increasing the flexibility of the teacher so that he can not only implement but also supplement existing curricula. Further research is needed in order to check on the findings of this study with different and larger samples as well as for different levels of the curriculum. The preservation of the videotape recordings of classroom behaviors may make possible further refinement of the observational instruments and hypotheses used in this study.

Table of Contents

Page

Introduction 1

Description of the Study 4

  Principal Hypotheses 4

  Selection of Teachers and Classes 9

  Variables Used: Teacher Behaviors 10

  Variables Used: Teacher Characteristics 18

  Variables Used: Student Attitudes 22

  Variables Used: Student Achievement 24

Computation 39

Results and Discussion 40

  Principal Hypotheses 40

  Correlational Analysis 51

Summary and Implications 69

  Development 69

  Research 70

References 73

Appendices 75

List of Tables

Table — Page

1. Teacher Characteristics as Indicated in Questionnaire 20

2. Raw Scores, MLA-ETS Language Proficiency Tests . 21

3. Student Attitudes toward Foreign Language Study: Class Means 23

4. Difficulty and Discrimination Levels of Listening Test 26

5. Difficulty and Discrimination Levels of Reading Test 27

6. Difficulty and Discrimination Levels of Grammar Test 28

7. Rater Agreement on Randomly Selected Writing and Speaking Tests 33

8. Information Concerning Language Laboratory Facilities and Use 38

9. t-Tests for Significance of Difference of Means for Classes Ranked as Top 5 and Bottom 5 on Teacher Behavior Variable 9 41

10. t-Tests for Significance of Difference of Means for Classes Ranked as Top 5 and Bottom 5 on Teacher Behavior Variable 10 45

11. Variables Possibly Affecting Achievement of Students in Classes Considered in Variable 10 48

12. Correlations of Criterion Measures of Student Achievement and Teacher Behaviors 63

13. Correlations of Criterion Measures of Student Achievement and Teacher Characteristics . . . 65

14. Correlations of Criterion Measures of Student Achievement and Student Attitudes 67

List of Appendices

Appendix Page

I. One-Way Analysis of Variance for Rater Reliability 75

II. Videotape Observation Rating Sheet 76

III. Ratings of Teacher Behaviors 77

IV. Class Means, Standard Deviations, Maximum, Minimum 78

V. Adjusted Class Means on Criterion Tests in Rank Order 80

VI. Teacher Questionnaire 81

VII. Mary Dufort Student Attitude Scale 84

VIII. Criterion Tests of Student Achievement 87

IX. Correlation Matrices 94

CHARACTERISTICS AND BEHAVIORS OF THE SUCCESSFUL FOREIGN LANGUAGE TEACHER

Robert L. Politzer and Louis Weiss*

Introduction

Efforts to improve the teaching of foreign languages have generally taken two avenues of approach: (1) the development of new teaching materials and new methodologies, and (2) the development of new and improved training programs for language teachers. With the advent of the National Defense Education Act in 1958, these two avenues converged into a broad effort toward excellence in foreign language teaching. NDEA Foreign Language Institutes in the United States and abroad took as their goal the retraining of foreign language teachers in the theory and practice of the so-called New Key approach, and new "audio-lingual" textbooks were adopted by school districts throughout the country.

It was also assumed by the NDEA teacher training programs that certain skills and knowledge were the necessary equipment of the foreign language teacher. Standardized tests were developed by the Modern Language Association to measure the teacher's knowledge and skills in the critical areas of Methodology, Applied Linguistics, Culture and Civilization, and Comprehension, Speaking, Reading and Writing of the language being taught. These tests formed the basis for the curriculum of the NDEA Foreign Language Institutes.

---

*Robert L. Politzer is Professor of Education and Romance Linguistics, Stanford University. Louis Weiss is a doctoral candidate in education and a Research Assistant at the Stanford Center for Research and Development in Teaching.

At the same time, however, very little research has studied the relationship between the teacher's knowledge and skills and student achievement. Ultimately, the value of teacher training programs must be judged in terms of effectiveness in the classroom. The task of the researcher, then, is one of identification and differentiation; the successful teacher must be identified by the achievement of his students on certain criterion measures, and the successful teacher must then be differentiated from the less successful one by specific types of classroom behavior and by certain personal characteristics. Research must assume that these specific behaviors and characteristics can be isolated and identified. Teacher training can then concern itself with the technique of teaching these specific behaviors to prospective teachers.

The main purpose of this study, therefore, was to observe as closely as possible the classroom behaviors of a group of teachers, to identify those who were successful in terms of pupil achievement, and to compare the behaviors and characteristics of the successful teachers with those of the teachers who were identified as less successful. The ultimate goal of the investigation was to determine whether certain assumptions made at present about successful classroom behaviors are valid and whether it is possible to identify specific teachable behaviors of successful language teachers.

Research in the practice of teaching in general and into the evaluation of teaching efficiency through pupil achievement has a long history (e.g., Gage 1963). In general, however, efforts to pinpoint successful classroom behaviors have been unsuccessful largely because of the "criterion problem," i.e., the question of how efficiency of teaching should be measured. It is surprising, therefore, that in the area of foreign language teaching, where student control of certain language learning skills provides observable and measurable criteria, so little effort has been made to investigate the nature of successful teaching behaviors. Description of the observation of good teaching practices has been undertaken (Brisley, et al. 1961) but generally without statement as to the criteria involved. A project of major proportions has been undertaken by McGill University in collaboration with the Center for Applied Linguistics (Hayes, Lambert and Tucker 1968) in which student achievement is used to measure the effectiveness of policies, principles, and classroom behaviors of language teachers.

Recent studies have investigated the effectiveness of the "audio-lingual" method as opposed to the "grammar-translation" method (Smith and Berger 1968; Smith and Baranyi 1968; Scherer and Wertheimer 1964), with ambiguous results. These studies attempted to transform the teacher variable to a constant. In the Smith and Berger study, for instance, certain behaviors thought to be typical of each of the "methods" were prescribed and were expected to be adhered to. However, to eliminate the teacher variable in this way may be quite unrealistic and artificial and may make it difficult to apply the findings of such a study to the real classroom situation. One must also ask if it is really the method that makes the difference in student achievement rather than the way in which the individual teacher implements that method with respect to his own competence, knowledge and characteristics as they interact with those of his students.

Comparison of methods is also made difficult by the problem of criterion measures, which must necessarily reflect the goals of teaching. The goals of the two methods compared in the Smith and Berger study are indisputably different. Thus, the question arises as to how to construct criterion measures that are valid for both methods, especially when these methods and measures are applied at the introductory level of language learning. It is not surprising that the "audio-lingual" students performed inadequately on tests designed to measure translation and reading ability and knowledge of rules of grammar.

The main independent variables of the present study are not methods but teaching behaviors and teacher characteristics. The method, not the teaching behavior, is a constant. Keeping the method constant made it possible to construct a set of criterion measures congruent with the goals of the textbook.

Certain contingencies of sampling procedure determined the selection of the textbook, the language being taught, and the level of instruction. In order to eliminate the effects of previous language experience, it was considered preferable to observe beginning classes in which students would be as close to the "zero point" as possible. In the particular geographical area in which this study was conducted, almost all students had been exposed to instruction in Spanish in the elementary school. French, as the only other language with an enrollment high enough to yield a satisfactory sample, was selected. A survey of textbooks revealed that A-LM French, Level One (Harcourt, Brace and World, 1961) was the only one used widely enough to provide a sufficient number of participating classes. Thus, it was decided to investigate the characteristics and classroom behaviors of 20 teachers of first-year French using A-LM as a textbook.

Description of the Study

Principal Hypotheses

There are at least two avenues of approach to the measurement of teaching efficiency. One consists in evaluating differential pupil achievement over certain time intervals. The other method has been called the micro-criteria approach (Gage 1968). Instead of the overall effectiveness of the teacher, the effectiveness of a specific behavior or behavioral skill (e.g., achieving closure, using questions, ability to explain) is evaluated with reference to a specific short teaching task. The latter method, which has been applied in various experiments at the Stanford Center for Research and Development (see First Annual Report of the Center, 1967; Gage, et al. 1969) has many advantages. The "micro-paradigm" makes it possible to focus on specific criterion measures and eliminates many uncontrolled variables. From the point of view of foreign language education research, it has the disadvantages that it is very difficult to arrange for situations in which numbers of students with similar or identical backgrounds and equal achievement in a language skill are taught the same short lesson by different teachers. In addition, it is, of course, necessary that the results derived from the micro-criteria approach be in turn applied to evaluation of global experimental instruction in order to determine their validity in a real-life, non-experimental situation. The proposed experiment, therefore, takes the real-life, global situation as its setting.

The study was designed to answer the following question: Which of certain selected teacher characteristics and classroom behaviors could be said to contribute to a significant difference in achievement of first-year foreign language high school students?

By means of specially constructed criterion tests, student and teacher questionnaires, MLA-ETS Proficiency tests, and observation of videotapes of classroom procedures, data were gathered that yielded 40 variables, as follows:

1-14	Teacher Behaviors
15-24	Teacher Characteristics
25-33	Student Attitudes
34-40	Criterion Tests of Student Achievement

In addition to an analysis of the intercorrelation matrices, two research hypotheses were formulated, one focusing on a linguistically definable aspect of the foreign language teacher's classroom behavior and the other on a pedagogical aspect.

The linguistic hypothesis. The procedure of the audio-lingual method and of the specific textbook used in this study requires the memorization of dialogues, the learning of grammar through carefully controlled structure drills and the practice of the language through directed dialogue and recombination of previously learned material. In one form or another, drill activity is central to the language learning procedure. For this reason, the behaviors selected for observation are those behaviors concerned with drill procedure. For purposes of analysis and discussion, the selected drills were assigned to two distinct categories, as follows:

(a) Controlled drills. This group is comprised of Repetition, Substitution, Translation, and Dialogue drills. These drills share one common feature: They do not require any manipulation of linguistic structure on the part of the student. In Repetition drills, the response is, of course, identical with the stimulus. In a pure Substitution drill, at least the kind in which the substitution does not require some sort of conversion, the structure of the response is identical with the structure of the model. Translation drill (or, rather, English cueing) and Dialogue drill require the verbatim recall of a construction.

(b) Free drills. This group contains two types of drills: Conversion drills and Free Response drills.

Conversion drills include the following drill activities:

Directed Dialogue, in which the student is told what to say in very specific terms (e.g., Cue: "Ask your friend how old he is." Response: "How old are you?")

Question-and-answer drills, in which the student's response is almost entirely suggested by the question (e.g., Cue: "Do you prefer rice or potatoes?" Response: "I prefer rice.")

The Free Response type of drill also includes question-and-answer drills, but of the type in which the student's response is not suggested by the stimulus (e.g., Cue: "What would you like with your steak?" Response: "I would like potatoes.")

Both the Conversion drills and the Free Response drills share one common characteristic: The linguistic structure of the response is not identical with that of the stimulus. It is for this reason that they have been grouped together into the Free drill category.

It has often been said that students who are restricted to Controlled drills and are not given enough exposure to Free drills will be able to do nothing more than repeat dialogues and drills in parrot-like fashion and will fail to comprehend or generate utterances other than the predictable ones they have memorized and drilled. On the other hand, students who have been exposed to a balanced proportion of both types of drill should be able to handle the language on both the predictable and non-predictable level. This was, in fact, one of the assumptions of the Performance Criteria for the Foreign Language Teacher, recently developed by Robert L. Politzer (Stanford Center for Research and Development in Teaching, Technical Report #1), in which it was suggested that "the final step in pattern practice should be the use of the pattern in response to a 'conversational' cue which is completely dissimilar to the pattern itself. . ." (p. 16).

The linguistic hypothesis may thus be stated as follows:

Hypothesis 1: The classes of the five teachers who rank highest on the proportion of free drill to controlled drill will perform significantly better on the achievement criterion tests than the classes of those five teachers who rank lowest on the same proportion.

The pedagogical hypothesis. The assumption underlying this hypothesis has also been expressed in the Performance Criteria mentioned above. "The recall and use of the pattern are, of course, also more likely to occur if the pattern is associated with or can be derived through a variety of processes rather than just one" (p. 16). The teacher who is sensitive to the effects of fatigue or boredom resulting from excessive drill of a particular type should be able to vary the drill activity by switching from one type of drill to another. In her discussion of the effects of fatigue, Wilga Rivers notes that "Homogeneity of task may cause fatigue. If students are kept too long at rote memorization of foreign language material, at drills (even of various types), and at repetition of dialogue material . . . then the task will become monotonous and actual learning will decrease. . . . The teacher should vary activities during the class lesson and develop sensitivity to class reaction. . ." (Rivers 1964, p. 69).

The pedagogical hypothesis, therefore, dealt not with the total proportions of one type of drill as opposed to another, but rather with the frequency with which the teacher switched from one type of drill activity to another. As will be discussed more fully in the detailed description of all the variables studied, the observational instrument used in the study recorded all the various types of drills used by a teacher within each minute of observed time. The minute-intervals during which the teacher engaged in both free and controlled type of drill were counted, and so were the minute-intervals during which the teacher engaged exclusively in either controlled or free drill activity. The ratio of the number of drill minutes of varied activity over the number of drill minutes of homogeneous activity was computed in order to serve as an index of how often the teacher changes from one type of drill to another. It will be noted that the instrument used in the study and the index of flexibility based on it (proportion of Switched drill to Exclusive drill; see Variable 10 below) did not make it possible to count more than one switch occurring within each minute of observation, or to take into consideration switches which happened to coincide exactly with the time boundary between the minute segments. A spot check on the possible number of such switches, undertaken by observing several of the videotapes, showed that the number of switches which were not counted because they were either the second switch within a minute or occurred at the minute boundary was for all practical purposes negligible.

The pedagogical hypothesis was thus stated as follows:

Hypothesis 2: The classes of the five teachers who rank highest in the proportion of minutes spent on both free and controlled drills over minutes spent homogeneously on only one type of drill will perform significantly better on criterion achievement measures than the classes of the five teachers who rank lowest on the same proportion.

Selection of Teachers and Classes

The teachers participating in the study were recruited from the staffs of three large school districts in the vicinity of San Jose, Calif. The general population in this area covers a broad socio-economic spectrum from auto and aircraft factory workers to professional workers in electronics and communications systems firms. A survey of the student population in the participating schools showed a range of college aspiration from 93% to 30%.

Three school districts in which A-LM Level One was used as a textbook in first year high school French were asked to take part in the study. The teachers using A-LM in their first year classes were invited to participate. None of those invited declined the invitation. Although the teachers were informed of the intent of the study, they were not aware of the main hypotheses. They were asked not to change their lesson plans or alter their teaching styles in any way during the videotaping sessions and were not informed that only their drill activities were to be observed and rated. During the year in which the study was conducted, all but one of the teachers took the MLA-ETS Language Proficiency Test. Teachers ranged in age from 51 years to 22 years, while the range of experience in teaching French was from 20 years to one year. None of the teachers chosen to participate was a native speaker of French.

Originally, 20 classes with a total of 512 students were selected to participate in the study, but the assignment of an excessive number of students repeating first-year French to three of the classes reduced the number of classes to 17. Although all students in the remaining 17 classes participated in all aspects of the study, those students who had studied French previously were not included in any of the statistical analyses. During the first week of school, all classes were given the Carroll-Sapon Modern Language Aptitude Test. Since the scores on this test were to be used as the covariate in adjusting the mean scores of the criterion tests, students who did not take the MLAT (because of absence or later assignment to the classes) were also dropped from the statistical analyses. The original number of subjects was thus reduced from 512 to 320.

Variables Used: Teacher Behaviors

Observational instrument for teacher behaviors.

The data relevant to this study were gathered through observation of videotapes recorded in the classrooms of the participating teachers. Ideally, such observation should cover the typical class period of a typical day in the teacher's career. Such an ideal would be attainable if a hidden camera were to record continuously all activities and if we were then able to determine what, in fact, was typical. To a certain extent, limitations of time and budget were determining factors in the following recording schedule:

Four 15-minute videotape recordings were made of each participating class, in the following sequence:

1. The first 15 minutes of the class period were recorded for all participating teachers during the week of October 3 to October 11.

2. During the period from October 31 to November 8, the recording was begun after the class had been in session for ten minutes.

3. During the week of November 28 to December 6, the recording was begun after the class had been in session for 20 minutes. This time segment was recorded early enough in December to avoid whatever effects the impending Christmas vacation would have had on the classroom activities.

4. From January 9 to January 17 the final 15 minutes of class were recorded.

Since participating classes were scheduled at different times of the day, it was possible to arrange a schedule to accommodate three or four classes each day. All the schools were close enough geographically to permit the operator to move easily from class to class.

An experienced videotape technician was instructed in the procedural details necessary to insure consistency and unobtrusiveness. All recordings were made by this operator. The recorder used was a Sony EV-200 for two-inch videotape. The recorder and camera are mounted permanently on a compact 30" x 20" x 44" cart mounted on 3" rubber casters. The equipment can thus be wheeled easily from classroom to panel truck and vice versa. In the interval between classes, the operator was able to set up the equipment at one side of the room in such a way as to permit observation of the interaction of teacher and students. This was facilitated by the use of a wide-angle lens attached to the camera. Two microphones were set up in the classroom, one in the center and the other at the rear. Each teacher was equipped with a Vega wireless microphone, affording complete freedom of movement. It was thus possible to see and hear everything that went on in the classroom during the observation periods.

At the beginning of each observation period, the operator turned the recorder on and kept it ready for use. At the second, third, and fourth sessions, the record mechanism was set in motion at the proper time. Although the teachers were informed of the time segments that were being recorded, the students were unaware of when or for how long the camera was actually in operation. Since the schools participating in the study were located in the general vicinity of Stanford University, most of the students and teachers were familiar with videotape observations (which are an integral part of Stanford's teacher training program) and were not especially distracted by the presence of the equipment or the operator.

At the end of the recording schedule, each teacher was represented by four 15-minute samples covering four months of the teaching year. The four-week interval was designed to minimize differences in teachers' behavior patterns during the semester. Some teachers may start off slowly and intensify their work in December and January, while others may start off vigorously and slow down as winter vacation approaches. Thus, if the drill activities relevant to this study did not show up in one 15-minute segment, they would most likely show up in another.

At the same time, however, no claim can be made that the 15-minute segments are an adequate and completely reliable sample of the total behavior of each teacher with respect to specific behaviors observed by the rating instrument. As we have already stated, the activities of the language classroom vary naturally during any one class period and throughout the year. The classroom activities in January tend to favor slightly different drill behaviors than those of December or October. Testing the reliability of the sample by comparing the results obtained from observing the same teacher during different time periods would therefore have only limited value. Thus, the sampling method necessitated by time and budget must be recognized as one of the limitations of the study. It was primarily for this reason that the conclusions of the study were not based on the research hypotheses alone, but also on a correlational analysis which made it possible to consider not just the effects of two independent variables but the relation of an entire set of variables to the criterion measures.

Rating of videotapes. All videotapes were rated by three teachers who were experienced in the use of A-LM but who were not associated in any way with the schools participating in the study. Prior to the actual observation sessions, the raters were given intensive training in the use of the rating instrument. Each behavior to be observed and noted was explained and discussed fully. In three separate sessions, the raters, along with the two investigators, observed and rated videotapes of former Stanford interns teaching first year A-LM French. These ratings were compared and discussed until all raters were in agreement on the definitions of the categories and the identification of the behaviors. During the third training session, three tapes were rated without comment or discussion. Rater agreement was then measured and proved to be high enough to warrant the beginning of the actual observation sessions.

At the first session, all three raters observed and rated four tapes as a group. These ratings were used to establish a level of rater agreement on four observations, namely: (1) frequency of Controlled drills per drill minute, (2) frequency of Free drills per drill minute, (3) number of minutes in which drills were exclusively Controlled or exclusively Free, and (4) number of minutes in which the teacher switched from Free to Controlled drills and vice versa. Rater agreement was perfect on measures (2) and (4) and a one-way analysis of variance showed an agreement of .97 and .99 on the other two measures (see Appendix I). Consequently, each of the remaining tapes was rated by only one of the raters working privately. No more than six tapes were observed at any one session and tapes were assigned in such a way that no rater observed the same teacher more than once at a single session. The raters were not familiar with any of the teachers whose tapes they rated and were not aware of the ranking of the classes in the criterion measures. As further insurance against rater bias, the hypotheses of the study were not made known to the raters.

Rating instrument for teacher behaviors. The instrument used in the rating of teacher classroom behaviors was designed to record objectively on three dimensions: Categories, Time, and Frequency. The raw data thus obtained could be expressed in terms of proportions or ratios. (A sample of the rating instrument may be found in Appendix II).

(a) Categories: Six types of drill were specified for observation and rating, two representing Free drills and four representing Controlled drills. On the rating sheet, the Controlled drills were labeled: (1) Repetition drill, (2) Substitution drill, (5) Dialogue drill and (6) Translation drill. (These drills were placed fifth and sixth for convenience of rating, since it was expected that their frequency of occurrence would be low.) The Free drills were represented by (3) Conversion drill and (4) Free Response.

(b) Time: The rating sheet contained 15 columns, one for each minute of observed time. During the rating sessions, an audio tape recorder was set up in each of the rooms and synchronized with the videotape recorder. At the end of each 60-second interval, a voice on the tape spoke the number of the following minute. This served as a signal for the rater to move on to the next column and also removed any question as to which one of the 15 minutes was being rated.

(c) Frequency: The rater recorded a tally for each drill event, that is, for each pair of stimulus/response involving teacher and students. An exchange was considered an "event" whether it was between teacher and individual or teacher and group. If there was no response from the student or if there was a series of exchanges between or among students, no tally was made. (The allowance for student to student interaction is discussed below.) From these tallies, it was possible to identify the variety and types of activities taking place within a 60-second interval and to obtain a proportion of total drill time devoted to each of the specified drills.

(d) General impression: In addition to their objective rating of drill activities, raters were asked to put down, at the end of each 15-minute observation, their general impression of other classroom behaviors. There were three columns labeled: Never, Occasionally, and Frequently. The cut-off point between Occasionally and Frequently was established for each behavior during the raters' training sessions.

During the training sessions, the raters agreed completely on these variables. The four categories of classroom behavior were:

1. Direct reference to book. How often did the teacher refer to the textbook or to cue cards during the drill activities? Five times or less was considered "Occasionally"; more than 5 times was considered "Frequently."

2. Use of visual aids. How often did the teacher use visual aids such as pictures, realia, overhead projector, chalkboard and so on during the drill activities? Ten was considered the cut-off point between "Frequently" and "Occasionally."

3. Student/student interaction. How often during the drill period did the teacher arrange for communication in French between students rather than between teacher and students? This category subsumed the use of the A-LM dialogues as well as free exchanges between students. Four or more occasions given to students to interact was considered as "Frequent."

4. Variation of structures. How often during the drill period did the teacher change the structures that were being

drilled? Although a variety of structures is provided by the textbook, it is quite possible for a teacher to drill extensively on only one structure to the point of fatigue and boredom. It was therefore considered important to take note of the number of grammatical problems the teacher treated during the observed time. The cut-off point for this measure was 4.

Variables 1-14: Block A. The raw data collected from the rating sheets were converted to proportions or averages to form the basis for the following fourteen variables of observed classroom behaviors. (For the table of scores for the 17 teachers, see Appendix III.)

Variable 1: Proportion of Time Spent in Drill Activity. A count was made of the number of columns in which there were any tallies (regardless of category) in the four observation sheets for each teacher. The total was divided by sixty to arrive at the proportionate amount of time each teacher spent drilling during the four observations. This in itself tells us little except the teacher's propensity for conducting drills. It does not tell us about the speed or type of drill conducted.

Variable 2: Average Number of Drills per Drill Minute (pdm). This variable tells us something about the speed with which the teacher conducts his drills but nothing about the type of drills he conducts. This figure was arrived at by adding all the tallies in all the cells and dividing the sum by the number of minutes during which the teacher drilled throughout the four observations.

Variable 3: Average Repetition Drills pdm. This figure was arrived at by dividing the total number of repetition drills for each teacher by the number of minutes the teacher drilled during the 60 minutes observed. The same process was used for variables 4, 5, 6, 7, and 8 that follow.

Variable 4: Average Substitution Drills pdm.

Variable 5: Average Dialogue Drills pdm.

Variable 6: Average Translation Drills pdm.

Variable 7: Average Conversion Drills pdm.

Variable 8: Average Free Responses pdm. The data gathered for this variable were disappointingly meager. Only seven of the teachers allowed for any free response at all and only one of these was of slightly more than negligible frequency. Teacher 2 allowed 18 free responses out of 48 minutes of drill; teacher 5 allowed 4 responses out of 38 minutes of drill; teacher 8 allowed 8 free responses out of 44 minutes of drill; teacher 2 allowed 2 responses out of 56 minutes of drill and the rest allowed one response each. Nevertheless, this variable was included in the numerator of the following variable.

Variable 9: Ratio of Free Drills to Controlled Drills. This ratio is the independent variable in the first hypothesis and is an indication of the relationship between the carefully controlled drill activities and the amount of freedom of expression afforded the students during the observed drill activities. It is arrived at by dividing the tallies for variables 3 (Repetition drills), 4 (Substitution drills), 5 (Dialogue drills), and 6 (Translation drills) into the total number of tallies for variables 7 (Conversion drills) and 8 (Free Response).

Variable 10: Proportion of Switched Drill to Exclusive Drill. This is the independent variable of the second hypothesis and is an indication of the frequency with which the teacher switched from Controlled drill to Free drill and vice versa. The figure was obtained by dividing the number of minutes during which the teacher worked exclusively in either Controlled or Free drill activities into the number of minutes during which the teacher switched from one type of drill to the other.

Variable 11: Reference to Book. For this variable as well as for variables 12, 13, and 14 that follow, a numerical score was given to the general impressions of frequency of occurrence. A score of 2 was assigned to "Never," a score of 4 was assigned to "Occasionally," and a score of 6 was assigned to "Frequently." The scores for the four observations were totalled and divided by four to arrive at the average score. In one case (Teacher 1) it was often impossible to see the teacher and the score of zero in this case represents "not observed."

Variable 12: Use of Visual Aids. This variable measures the frequency with which the teacher used pictures, realia, overhead projector, and chalkboard during the drill activities.

Variable 13: Student to Student Interaction. This variable measures the amount of student to student interaction in French as part of the drill activities and includes the recitation of dialogues as well as relatively free interchange between students.

Variable 14: Variation of Structures. This variable measures the variety of structures (grammatical problems) the teacher introduced within a 15-minute segment.

Variables Used: Teacher Characteristics

Data concerning the teacher characteristics variables used in the correlation matrix were gathered from two sources, the MLA-ETS Proficiency Tests and a specially prepared questionnaire.

Questionnaire. Toward the end of the school year, all teachers were asked to complete a questionnaire (see Appendix VI). In addition to information concerning sex, age, experience, college major, and time spent in France or a French-speaking country, the teachers were asked to give their opinion of the textbook by placing a checkmark in one of three columns on a semantic differential scale of ten pairs of adjectives. A checkmark in the positive column was valued at 3 points, a checkmark in the neutral column was valued at 2 points, and a mark in the negative column was valued at 1 point. The points were tallied to give a global rating with a maximum value of 30 and a minimum value of 10. The average rating for all teachers was 19.76.

Variables 15-24: Block B. Table 1 below includes pertinent information gathered from the questionnaire. Columns marked with an asterisk indicate variables used in the correlation matrix. These are:

Variable 15: _Teacher's Attitude toward the Textbook_.

Variable 16: _Years of Experience in Teaching French_.

Variable 17: _Number of Months Spent in French-Speaking Country_.

MLA-ETS Foreign Language Proficiency Tests. In spite of the fact that some of the teachers had taken the MLA-ETS tests previously, all teachers were asked to take the tests during the year of the study. Of the 17 teachers participating in the study, only one (Teacher 14) was unable to take the tests.

Scores on these tests are represented by the following variables in the correlational matrix:

Variable 18: _Listening_

Variable 19: _Speaking_

Variable 20: _Reading_

Variable 21: _Writing_

Variable 22: _Applied Linguistics_

Variable 23: _Culture and Civilization_

Variable 24: _Professional Preparation_

Table 2 shows the scores for each teacher on each of the tests, the means and standard deviations of the 16 participating teachers and the means and standard deviations obtained for the post-tests taken by participants in NDEA Foreign Language Institutes from 1961 to 1965.

TABLE 1

Teacher Characteristics as Indicated in Questionnaire

Teacher number	Sex	Age	Months in French-speaking country*	Under-graduate major	NDEA Institute	Years teaching*	Attitude to text*
1	F	24	12	French	No	2	25
2	F	26	36	French	No	2	19
3	F	31	0	Art	Level I	7	20
4	F	25	3	French	No	2	23
5	F	22	0	French	No	1	13
6	M	37	3	Spanish	Level I & II	8	21
7	M	34	6	Psych	No	5	26
8	F	22	1	French	No	1	12
9	F	51	12	French	No	20	29
10	F	25	12	French	No	3	17
11	F	25	1	Spanish	No	1.5	25
12	F	23	12	French	No	2	21
13	F	?	3	Speech	No	7	14
14	F	?	6	French	No	1	13
15	F	24	6	French	No	2	16
16	M	30	3	French	Level I & II	8	17
17	F	26	12	French	No	1	25

Note: Columns marked with an asterisk indicate variables used in the correlation matrix.

TABLE 2

Raw Scores, MLA-ETS Language Proficiency Tests

Teacher	List.	Spk.	Rdg.	Wrtg.	Ling.	C & C	Prof. Pr.
1	49	84	54	59	53	46	60
2	58	81	57	61	62	54	67
3	43	75	39	50	55	48	76
4	39	69	39	45	52	46	65
5	48	74	47	49	49	44	65
6	48	78	60	57	69	70	71
7	29	62	30	31	49	39	66
8	43	76	52	58	61	53	67
9	43	64	41	39	34	43	59
10	51	76	70	65	61	57	76
11	51	65	41	47	53	54	73
12	43	73	49	43	50	46	62
13	47	99	64	67	61	57	76
14							
15	56	85	60	57	62	57	72
16	45	89	53	40	61	56	71
17	52	88	49	54	58	44	61
Means:	46.6	77.4	50.3	51.4	55.6	50.9	67.9
s.d.:	6.9	10.0	10.6	10.1	8.1	7.8	5.8
NDEA Means:	42.8	80.2	45.3	45.2	51.2	49.1	67.0
s.d.:	8.4	16.1	10.3	12.3	8.8	8.4	6.5

Variables Used: Student Attitudes

For the measurement of student attitudes toward the study of Foreign Languages, the instrument first developed by Dr. Mary Dufort in 1962 was used. (See Appendix VII.) This instrument is a Likert-type scale developed to measure the attitude of sixth grade students toward Foreign Language Instruction. The scale has a test/retest reliability of .87 and in an experiment conducted by Mrs. Diana Bartley on the Junior High School level, the scale was found to have a reliability of .92 by the Cronbach Alpha formula (see Cronbach 1951). Some very slight revisions were made in the wording of the instrument in order to make it adaptable to the High School level. The Mary Dufort Foreign Language Attitude Scale consists of thirty items reflecting a positive attitude toward the study of the language, the speakers of the language, and so on. Students were asked to note whether they agreed (1) not at all, (2) a little bit, (3) quite a bit, and (4) very much. Each item was scored on a basis of 1 to 4 points on an ascending scale. The maximum score on the scale is thus 120 and the minimum score is 30. Two items were added to the original 30. In these items, the students were asked to rate the statements: (31) "I would like to continue my study of French next year," and (32) "French is one of the most important subjects in the curriculum."

Variables 25-33: Block C. Since it was the intent in this study to measure the gain or loss in positive attitude during the time the students were with the participating teachers, the scale was administered twice, the first administration being during the first week of school along with the MLAT, the second being in May, just before the end of the school year. The results of these two administrations and the differences between them provided Variables 25 to 33, as shown in Table 3.

TABLE 3

Student Attitudes toward Foreign Language Study: Class Means

	General Attitude Measures			Intent to Continue			Importance of Subject		
Teacher	Pre-test Var. 25	Post-test Var. 26	Gain Var. 27	Pre-test Var. 28	Post-test Var. 29	Gain Var. 30	Pre-test Var. 31	Post-test Var. 32	Gain Var. 33
1	78	84	6	3.24	3.22	-.02	1.65	1.44	-.21
2	77	80	3	3.19	3.00	-.19	1.69	1.57	-.12
3	71	*		2.91	*		1.91	*	
4	80	74	- 6	3.22	2.74	-.48	1.72	1.79	.07
5	91	88	- 3	3.56	3.22	-.34	2.63	2.15	-.48
6	80	77	- 3	3.35	3.50	.15	1.61	1.73	.12
7	82	67	-15	2.53	2.53	0	2.26	1.82	-.44
8	79	75	- 4	3.22	2.76	-.46	2.06	1.64	-.42
9	77	74	- 3	2.91	2.75	-.16	2.00	1.80	-.20
10	77	81	4	2.88	2.87	-.01	1.92	2.00	.08
11	78	74	- 4	3.04	2.74	-.30	1.89	1.61	-.28
12	80	79	- 1	3.07	3.00	-.07	1.93	1.71	-.22
13	74	71	- 3	2.83	2.74	-.09	1.67	1.58	-.09
14	76	71	- 5	2.90	2.63	-.27	2.03	1.48	-.55
15	65	76	11	2.63	2.69	.08	1.58	1.75	.17
16	79	78	- 1	3.23	2.88	-.45	2.10	2.00	-.10
17	83	82	- 1	3.52	3.05	-.47	1.95	2.00	.05

*Post-test was not given to this class.

Variables Used: Student Achievement

Description of criterion tests. Since the conclusions reached in this study were based on the students' ability to perform certain tasks in the language, great care was taken in the construction of the criterion tests which were to measure student achievement at the end of the first semester. Construction of a special test was necessary since there is no standardized test available that is designed to be given at the end of the first five months of language study. Achievement after five months is so restricted to a specific curriculum that a standardized test designed to measure language achievement at this early stage would be almost impossible to construct. The participating teachers were polled to determine the number of A-LM units they expected to cover during the first semester as well as the reading and writing goals established for this period. All teachers planned to complete unit six and some planned to cover part of unit seven. The vocabulary and structures used in the criterion tests were therefore restricted to those appearing in the first six units of the textbook. None of the teachers felt that any tests, including those in reading, grammar, and writing, were inconsistent with their own goals.

A preliminary form of the test was constructed by the investigators, one of whom had had seven years experience in teaching A-LM in secondary schools as well as experience in teaching the Methods course at a second-level NDEA Institute. Two weeks before the actual administration of the test, this preliminary version was pre-tested in two first-year French classes in a school district not involved in the study. A divided sample, full item analysis (Valette 1967, p. 40) was made on all components of the criterion measure and the test was further modified. The revised criterion test (see Appendix VIII) was then given to all participating subjects, and consisted of seven subtests, as follows:

Variables 34-40: Block D

Variable 34: Listening. This test was similar in form to the A-LM Listening Tests and contained items that covered work done in the first six units. The test consisted of ten questions, each followed by three response choices. Of the two distractors for each item, one was based on a slight difference in meaning and the other on a slight difference in sound. The questions and responses were recorded on audio tape and the students were required to mark their choice on a specially prepared answer sheet. Each question was read twice and then followed by only one reading of the response choices. Students were informed of this and instructed to listen carefully. A pause of five seconds was provided to permit students to mark their answer sheets. Each item was given a value of one point, so that the maximum score on this subtest was ten. A formal item analysis using the upper and lower 27 per cent of the test papers was made to determine the level of difficulty and the discriminatory power of each item. Table 4 shows the results of this analysis, obtained according to the following formulas:

Level of difficulty

$$\text{Difficulty} = \frac{\text{Number of items correct in both groups}}{\text{Total number of items in both groups}}$$

The acceptable range is from .20 to .80 with the larger quotients representing less difficult items. It will be noted that only the first item falls outside the acceptable range.

Discriminatory power

$$\text{Discrimination} = \frac{\begin{pmatrix}\text{Correct items}\\ \text{upper 27\%}\end{pmatrix} - \begin{pmatrix}\text{Correct items}\\ \text{lower 27\%}\end{pmatrix}}{\text{27\% of total test items}}$$

In this case, the higher quotient represents greater discriminatory power. Quotients higher than .20 are considered acceptable. As will be noted in Table 4, the reliability quotient for the Listening Test was .84 as obtained with the Kuder-Richardson Formula #20 (Kuder and Richardson 1939).

TABLE 4

Difficulty and Discrimination Levels of Listening Test

Item	Difficulty	Discrimination
1	.86	.26
2	.70	.51
3	.75	.48
4	.62	.57
5	.58	.58
6	.37	.44
7	.72	.53
8	.53	.65
9	.57	.65
10	.47	.76

Reliability = .84

Variable 35: Reading. The purpose of this test was to measure the student's comprehension of a written paragraph, based on his ability to identify correct grammatical forms rather than rote mastery of certain sentences contained in the dialogues and drills. The student was presented with a short paragraph of 57 words, seven of which had been deleted. These seven words, along with seven distractor-words, were listed in random order in a column to the right of the paragraph. The student was to choose the correct word and write it in the proper space. Each correct word could be used only once and only in its proper blank space. Each item was given a value of one point, so that the maximum score on this test was seven. Table 5 shows the results of the item analysis made on this subtest.

TABLE 5

Difficulty and Discrimination Levels
of Reading Test

Item	Difficulty	Discrimination
1	.39	.68
2	.47	.74
3	.62	.63
4	.55	.69
5	.42	.52
6	.57	.74
7	.64	.69

Reliability = .85

Variable 36: Grammar. In this test, the student was required to rewrite five sentences from plural to singular form. The task involved several grammatical principles: (a) the singular/plural relationship of nouns and verbs, (b) the singular/plural relationship of nouns and modifiers, (c) the agreement in gender of nouns and their modifiers. Since gender is not marked in the plural form of the article, the task of changing a sentence from plural to singular requires the student to choose the correct form of the modifier. A value of one point was given to each verb, noun, and modifier, with the exception of "jeunes filles," which was given a value of two points. The total maximum score for this test was 16. Table 6 shows the results of the item analysis made on this subtest.

TABLE 6

Difficulty and Discrimination Levels of Grammar Test

Item	Difficulty	Discrimination
1	.40	.71
2	.84	.31
3	.52	.84
4	.76	.47
5	.81	.36
6	.39	.69
7	.59	.64
8	.82	.31
9	.53	.86
10	.24	.49
11	.67	.50
12	.22	.44
13	.63	.69
14	.80	.37
15	.83	.33
16	.45	.77

Reliability = .92

Variables 37 and 38: Base (37) and Free (38) Writing. A test was designed to measure the student's writing ability on two levels: Base and Supplemental. The student was presented with five questions, four of which were closely related to sentences found in the A-LM dialogues. The fifth question: "Pourquoi n'écoutes-tu pas tes disques?" was intended to provide an opportunity for the more proficient students to respond. Students were instructed to answer the questions to the best of their ability and were encouraged to write as freely and as much as possible by the following grading system:

Base score: Students were instructed to answer each question with a complete sentence. A maximum value of five points was given to this initial response with a deduction of one point for each error, not to exceed five. That is to say, a student could receive a zero for a completely incorrect response but never a minus score. Following are examples of acceptable base responses:

1. Où vas-tu?

   Je vais au cours de français. *Je vais à la bibliothèque*. Two points were deducted if the student omitted the subject and verb, so that an answer like: *Au cours de français* or *A la bibliothèque* was valued at only three points.

2. Où est-ce que tu habites?

   J'habite en face de l'église (école, etc.).

3. Qu'est-ce qu'il y a à manger aujourd'hui?

   Il y a des saucisses (des frites, du riz, etc.). Two points were deducted for omission of *il y a*. No credit was given for the use of *Je mange*, as this was not considered a response to *Qu'est-ce qu'il y a ...* .

4. Qu'est-ce que tu fais après l'école?

   Pas grand'chose was valued at only two points, but any other short, complete sentence (e.g., *Je vais chez moi*, *Je joue aux cartes*, etc.) was awarded full credit.

5. Pourquoi n'écoutes-tu pas tes disques?

   Le pick-up ne marche pas (the usual response to this question) was valued at five points.

In the analysis of the data, this part of the test is referred to as Base Writing (Variable 37). The maximum score on this part of the test was 25.

Supplemental score: For the sake of brevity, this part of the test is referred to as Free Writing (Variable 38). The grading system for this part of the test was carefully explained to the students in order to ensure as much free (that is, supplemental) writing as possible. Each additional clause or phrase that could be considered a "thought group" relevant to the question was given a value of five points. In the case of Question 3 ("Qu'est-ce qu'il y a à manger?"), each item of food correctly listed after the base response was valued at one point. Deductions were made for errors in spelling and grammar, but the deductions never exceeded the points given for each supplemental response. Our hopes that this grading system would encourage free and spontaneous writing were not groundless. A surprising number of students were eager and able to respond to the questions with more than just a basic sentence, in spite of the fact that they were only in their fifth month of language study.

Following are examples of responses and scores taken from actual test papers. Although there was no set maximum score for this part of the test, the actual maximum attained was 99.

1. Où vas-tu?

+5 +5 -1 -1
Je vais à la bibliotheque / pour etudie le français

+5 -1 -1 +5
lesson / ma proffessuer donne moi. / Je cherche mon papier /

+5 -1(je) -1 +5 -1 +5 -1
et regarder le livre aussi. / Alors je vais dejeune / avec mon

+5 -1 +5
amies / et apres dejeuner / je vais au cours de français.

2. Où est-ce que tu habites?

+5 +5
J'habite loin d'ici, / en face de l'école.

3. Qu'est-ce qu'il y a à manger aujourd'hui?

+5 +5 -1
il y a des saucisses. / Je n'aime pas des saucisses. /

+5 +1
Mais j'aime les frites et du lait.

4. Qu'est-ce que tu fais après l'école?

No credit--does not answer question   +5 -1(je donne)
On peut ne acheter du pain et du beurre. / Donnon un

+5
coup de telephone a Marie / pour regarder le television avec

+5   -1   +5
moi / et mange du pain et du beurre. / Alors, nous allons chez

+5   -1 (pour)
Anne / et ecouter les disques et danser.

5. Pourquoi n'écoutes-tu pas tes disques?

+5   +5
Je n'écoute pas les disques / parce que le pick-up ne

+5   -1   +5
marche pas. / Mais je peux aux cartes / avec Jeanne et Alice. /

+5   -1   +5
Alors, nous peutons aller en ville / et ecouter les disques.

Variables 39 and 40: *Base (39) and Free (40) Speaking*. This test was designed to measure the students' speaking ability on two levels: Base and Supplemental. The grading on this test was exactly the same as that of the writing test with pronunciation errors replacing spelling errors. Since many of the teachers used the "tu" form of address in their interaction with the students as well as in the drills, the questions were put in this form. The test consisted of twelve questions, recorded on audio tape with sufficient response time allowed between questions. The questions were of increasing difficulty and were grouped in the following way:

The first question ("Comment t'appelles-tu?") served to put the student at ease as well as to identify him by name.

The second, third, and fourth questions were simple arithmetic problems to which the student was instructed to respond by repeating the problem (e.g., Q: *Combien font deux et deux?*;

R: Deux et deux font quatre. Quatre alone was not considered a correct response but was given a value of one point.)

Questions 5 ("Quel âge as-tu?") and 6 ("Quel temps fait-il?") were included as examples of frequently asked questions requiring simple responses.

Questions 7, 8, and 9 offered the student a choice of two responses, both included in the questions themselves (e.g., "Tu habites loin d'ici ou près d'ici?"). The student was required to make the proper structural conversion in order to answer the questions. These questions paralleled the Conversion drills (Dialogue adaptation, etc.) carried on in the classroom setting.

The last three questions were designed to measure the amount of French the student could generate freely and spontaneously (e.g., "Pourquoi vas-tu à la bibliothèque après l'école?"). The following taped instructions preceded the questions:

> "In answering the last three questions, feel free to say whatever you like and as much as you like. You will be given extra time to make your response."

A pause of 45 seconds was allowed after each of the last three questions. The grading on this test was as follows:

Base score: A maximum of five points was given for the first complete sentence in response to each question, with a deduction of one point for each pronunciation or grammatical error, not to exceed five. In the analysis of the data, this part of the test is called: Base Speaking (Variable 39). The maximum score for this part of the test was 60.

Supplemental score: For the sake of brevity, this part of the test is referred to as Free Speaking (Variable 40). As in the Free Writing test, each additional clause or phrase that could be considered a "thought group" relevant to the question was given a value of five points, with deductions for errors in pronunciation and grammar not exceeding the points given for the response. Although only the last three questions were designed specifically for such supplemental responses, some students took the opportunity of expanding on their responses to questions 7, 8, and 9 and were given due credit. Although there was no set maximum score for this part of the test, the actual maximum attained was 19.

Grading procedures for Speaking and Writing tests: For the grading of the Speaking and the Writing tests, two French teachers who were thoroughly familiar with A-LM were instructed in the grading rules and procedures. One rater was assigned to the Speaking test and the other to the Writing test. In addition, each of the raters graded a randomly selected set of his opposite member's tests. These grades were used to test rater agreement. Table 7 shows the means, standard deviations, and correlation coefficients of the independently graded tests, indicating extremely high rater reliability. As can be seen by Table 7, the variation in performance on this test was considerable since the test was open-ended in the Free components.

TABLE 7

Rater Agreement on Randomly Selected
Writing and Speaking Tests

Test	Mean	sd	Mean	sd	r
Writing (n = 19)					
Base	13.47	6.69	12.47	6.94	.96
Free	21.89	22.37	19.31	18.64	.98
Speaking (n = 14)					
Base	44.64	11.93	44.07	12.03	.98
Free	9.57	11.79	8.50	10.54	.99

Administration of the criterion tests. To ensure uniformity of the testing procedure, the entire criterion test was administered to all classes by one of the investigators. By using the videotape schedule, it was possible to test all classes within a period of seven school days. The only exception to this was the Speaking test, which had to be completed during the following week. In general, however, we can say that the classes tested during the last days had not had enough extra instruction to give them any advantage over the classes tested during the first days.

On the test day, the classroom teacher turned the class over to one of the investigators, who gave identical instructions to all classes.

The Listening test was the first to be given. All classes heard the same tape, played on the same tape recorder (a Wollensack 1500), which was placed in the front of the room. During the Reading, Grammar, and Writing tests, the classroom teacher acted as proctor while the investigator administered the Speaking test. Due to the exigencies of time and money, it was considered impossible to give the Speaking test to all students. Thus, only fourteen were tested from each class, selected randomly within each sex group. The procedure for the Speaking test was the same for all classes. Two tape recorders were set up in one corner of the classroom, arranged so that the student taking the test was facing away from the rest of the class. The questions to the Speaking test were played on one tape recorder and received by the student through a head-set which had the advantage of cutting off any extraneous noise that might distract the listener. The student recorded his responses to the questions on a second tape recorder. The investigator operated both tape recorders and moved away from the student while he was taking the test, so as to minimize tension and embarrassment on the part of the student. The entire Speaking test took five minutes and those students who were tested on the first day were given an extra five minutes at the end of the testing period to complete the written part of the test. On the first day, four or five students in each class were able to take the Speaking test. The other nine or ten students were tested at the earliest convenient date. All testing was completed within three weeks.

Adjustment of criterion measures. Aptitude: As expected, most of the criterion measures correlated highly with individual student aptitude scores (MLAT). The specific correlations for the seven criterion measures were:

Listening (Variable 34)	.48
Reading (Variable 35)	.51
Grammar (Variable 36)	.56
Writing, Base (Variable 37)	.57
Writing, Free (Variable 38)	.49
Speaking, Base (Variable 39)	.54
Speaking, Free (Variable 40)	.26

In other words, the Free Speaking test, probably highly influenced by motivation, personality factors, and so on, had the lowest correlation with MLAT while higher correlations were reached by other criterion measures (the highest being with Writing and Grammar).

An examination of the correlations of MLAT scores and achievement by individual classes was made. This examination revealed that the relations between MLAT scores and criterion measures 34 (Listening) and 38 (Free Writing) were not the same for each class. Thus, adjustment for aptitude was not possible for these criterion measures. For the others (35, 36, 37, 39, 40), the mean scores of all classes were adjusted according to the following formula:

$$\text{Adjusted score} = \overline{Y}_i - b_p\,(\overline{X}_i - \overline{X}_c)$$

in which:

$\overline{Y}_i$ = raw score

b_p = regression slope for the pooled equation for the variable

$\bar{X}_i$ = MLAT mean score for the individual class

$\bar{X}_c$ = MLAT combined mean score.

(For relevant data on the MLAT, see Appendix IV.)

Attitude, sex, laboratory procedure, college orientation: In addition to aptitude and the teacher characteristics and behaviors under consideration in this study, we also considered student attitude, sex composition of the classes, college orientation, and the use of the language laboratory.

As will be discussed later (in connection with the correlational analysis), some of the attitude measures taken early in the course (Variables 25, 28, 31) do in fact show a tendency to correlate negatively with some of the mean scores of the criterion measures. In other words, the higher the initial attitude, the lower the eventual achievement. Evidently, there is a negative relation between initial high expectations and achievement. However, it was not felt advisable to adjust in any way for attitude scores, primarily because even initial attitude scores (obtained during the first ten days of instruction) may very well have been influenced by some of the teaching behaviors. This suspicion seems to be confirmed by the negative correlation (-.52) of initial attitude (Variable 25) with the frequency of the use of Repetition drills (Variable 3). (See Appendix IX-f, Block AC.)

In order to check the hypothesis that the sex composition of the classes may have had a significant relation to achievement even after adjustment for aptitude, the adjusted class means on all the criterion tests were correlated with the ratio of female/male in each class. The correlations for the seven criterion variables were:

Listening (Variable 34)	-.002
Reading (Variable 35)	-.22
Grammar (Variable 36)	-.03
Writing, Base (Variable 37)	.20
Writing, Free (Variable 38)	.10
Speaking, Base (Variable 39)	.05
Speaking, Free (Variable 40)	.16

None of the above correlation coefficients approached significance for an N of 17.

To assess the significance of the utilization of the language lab on the criterion measures was a rather difficult task, primarily because detailed data which would lend themselves to meaningful quantification were not really available. The exact type of lab activity is often rather difficult to determine and to describe. In general, the classes taking part in the study used only those tapes available in conjunction with the A-LM materials, i.e., the recordings of the dialogues and pattern drills.

The available information concerning the utilization of the language lab is summarized in Table 8. From it, it is difficult to discern any pattern of lab utilization which could be used to account for achievement differences. Since lab activity can be assumed to have its most immediate impact on listening and speaking skills, correlations between the mean scores on the Listening and Speaking tests (Variables 34 and 39) and the average number of minutes of lab work per week were established. For the Speaking test, the correlation is not significant (-.04), but the correlation for the Listening test (-.44) approaches significance (.1 level < .41, .05 level < .48). This may very well be an indication that the type of laboratory activity implied in dialogue memorization and pattern drill does little to increase auditory comprehension and may, if it is undertaken at the expense of other exercises, have a negative influence on it. (See Mathieu 1962 and Rivers 1964 for comments on the teaching of auditory comprehension skills.)

TABLE 8

Information Concerning Language Laboratory Facilities and Use

Teacher	Type of lab (Level)	Regular use	During or out of class	Times/ week	Min./ week	Type of program		
						Dial	Drill	Suppl.
1	III	No	During	1	10	x		
2	I	No	Out	1	30		x	
3	II	Yes	Out	2	50		x	
4	I	Yes	Out	2	50	x	x	
5	I	No	Out	2	30	x	x	x
6	II	Yes	Out	2	45		x	
7	II	No	During	1	35	x	x	
8	II	Yes	During	2	45	x	x	
9	I	Yes	Out	2	60		x	
10	II	No	During	1	20	x	x	
11	I	No	Out	2	30	x	x	x
12	III	No	During	1	10	x		
13	II	No	During	2	40	x	x	x
14	II	Yes	During	2	45	x	x	
15	II	Yes	During	2	45	x	x	
16	II	Yes	Out	2	50		x	
17	II	Yes	Out	4	80	x	x	

That the college orientation of the students may influence their motivation and achievement is a plausible hypothesis. In order to check the influence which the general school milieu may have had on the criterion scores, data were obtained concerning the percentage of college expectations of the student bodies of the participating schools. Again, the mean scores [illegible] the achievement tests were correlated with the percentage of college-oriented students in each of the schools. None of the correlations is significant:

Listening (Variable 34)	-.22
Reading (Variable 35)	-.15
Grammar (Variable 36)	-.07
Writing, Base (Variable 37)	-.06
Writing, Free (Variable 38)	.14
Speaking, Base (Variable 39)	.24
Speaking, Free (Variable 40)	.08

There was, therefore, no reason to assume that the percentage of college orientation within the student body had a significant relationship to the mean achievement scores according to which the classes were ranked.

Computation

All the data on the 40 variables considered in the study were coded and punched on IBM cards for processing by IBM 360/67 Computer. A program written by the Methodology Unit of the Stanford Center for Research and Development in Teaching (Program SCRDT 1) was used for an analysis of the regression of achievement scores on aptitude (MLAT) scores. From these data, it was possible to adjust the raw scores on five of the seven achievement tests. The adjusted scores (see Appendix V for adjusted class means in rank order) were then used to obtain a 40 x 40 correlation matrix using the Biomed Computer Program BMD - 02D which is a program for correlation with transgeneration (see Dixon 1967, pp. 49-59).

Results and Discussion

Principal Hypotheses

Hypothesis 1. The hypothesis stated that teachers having higher proportions of Free drill (Conversion and Free Response drill) to Controlled drill (Repetition, Substitution, Dialogue, Translation) would achieve better results than the teachers whose activities were dominated by the controlled type of drill. In order to test this hypothesis, the achievement criterion tests of the classes of the five teachers who ranked highest on Variable 9 (Ratio of Free drill to Controlled drill) were compared with those of the five teachers who ranked lowest on the same variable.

To test the significance of difference in achievement, the following procedure was used for each of the criterion measures (Variables 34 through 40): The class of each teacher who ranked in the high group on Variable 9 was paired randomly with the class of a teacher who ranked in the low group. For each pair, the following formula was used to obtain a measure of the difference of means between classes of unequal *n*:

$$d_i = \frac{\overline{X}_{i1} - \overline{X}_{i2}}{\sqrt{\frac{1}{n_{i1}} - \frac{1}{n_{i2}}}}$$

in which $\overline{X}_1$ and $\overline{X}_2$ represent the adjusted mean scores and n_1 and n_2, the number of students in each class. The means and standard deviations of d_i (difference of means between classes with unequal *n*) were calculated for the five matched pairs of classes and the following formula was used to obtain the t value of the difference between the means of the five pairs:

$$t = \frac{\overline{d}}{sd/\sqrt{5}}$$

As can be seen from Table 9 none of the t values for the differences in achievement on any of the criterion measures (Variables 34 through 40) reaches significance. Thus, Hypothesis 1 is not substantiated. The Ratio of Free to Controlled Drill (Variable 9), as observed in the sample of teaching behavior on which this study is based, does not differentiate the successful from the less successful teacher.

TABLE 9

t-Tests for Significance of Difference of Means
for Classes Ranked as Top 5 and Bottom 5
on Teacher Behavior Variable 9

(All differences are weighted differences
for classes of unequal number)

Class # (High)	Mean	n	Class # (Low)	Mean	n	d_i
Listening test						
14	6.62	24	3	4.00	17	8.29
2	7.24	25	6	5.68	25	5.53
7	6.31	16	13	5.24	17	3.07
8	5.17	12	12	6.30	10	-2.64
10	6.92	12	11	7.96	26	-2.98

The paired t-test based on $\bar{d}$ yields $t = 1.01$ with 4 df, ns.

TABLE 9 (Continued)

Class # (High)	Mean	n	Class # (Low)	Mean	n	d_i
Reading test						
14	3.80	24	3	3.40	17	1.27
2	3.92	25	6	3.28	25	2.27
7	2.96	16	13	3.62	17	-1.90
8	4.32	12	12	4.37	10	-0.12
10	3.69	12	11	4.91	26	-3.50
The paired t-test based on $\bar{d}$ yields t = -.38 with 4 df, ns.						
Grammar test						
14	10.68	24	3	7.62	17	9.75
2	10.44	25	6	8.37	25	7.34
7	8.05	16	13	9.20	17	-3.30
8	11.74	12	12	9.84	10	4.44
10	9.14	12	11	10.89	26	-5.01
The paired t-test based on $\bar{d}$ yields t = .91 with 4 df, ns.						
Writing test (Base)						
14	13.48	24	3	9.07	17	13.96
2	18.69	25	6	10.33	25	29.65
7	10.72	16	13	9.39	17	3.82
8	12.73	12	12	13.07	10	-.79
10	10.70	12	11	16.33	26	-16.13
The paired t-test based on $\bar{d}$ yields t = .80 with 4 df, ns.						

TABLE 9 (Continued)

Class # (High)	Mean	n	Class # (Low)	Mean	n	d_i
Writing test (Free)						
14	19.13	24	3	8.58	17	33.39
2	47.88	25	6	9.40	25	136.45
7	7.13	16	13	8.53	17	-4.02
8	17.92	12	12	11.00	10	16.17
10	14.67	12	11	27.81	26	-37.65
The paired t-test based on $\bar{d}$ yields t = .98 with 4 df, ns.						
Speaking test (Base)						
14	37.55	10	3	23.58	5	25.54
2	46.50	12	6	28.10	13	46.23
7	25.63	5	13	33.13	10	-13.71
8	41.06	6	12	25.04	7	28.92
10	31.03	8	11	40.43	12	-20.61
The paired t-test based on $\bar{d}$ yields t = 1.02 with 4 df, ns.						
Speaking test (Free)						
14	2.14	10	3	.14	5	2.14
2	9.72	12	6	1.53	13	9.72
7	-.42	5	13	1.03	10	-.42
8	2.78	6	12	1.61	7	2.78
10	2.39	8	11	1.83	12	2.39
The paired t-test based on $\bar{d}$ yields t = 1.96 with 4 df, ns.						

Hypothesis 2. The hypothesis stated that higher pupil achievement will be found in the classes of teachers who switch frequently from the controlled type to the free type of drill. In order to achieve a measure of this type of flexibility on the part of the teacher, the number of minutes during which the teacher switched from Controlled to Free drill (or vice versa) was divided by the number of minutes during which only one type of drill was used exclusively (Variable 10).

The statistical procedures used to test Hypothesis 2 were the same as those employed for Hypothesis 1: The classes of the five highest and five lowest ranking teachers on Variable 10 were compared for all the criterion achievement measures. As can be seen from Table 10, on five of the seven criterion measures (namely: 36, Grammar; 37, Base Writing; 38, Free Writing; 39, Base Speaking; 40, Free Speaking) the classes of the five highest-ranking teachers achieved significantly better than those of the five low-ranking teachers. For criterion measures 34 and 35 (Listening and Reading) the differences in achievement were also in favor of the high-ranking teachers, but did not reach significance. At any rate, Hypothesis 2 seems clearly sustained. Teachers who vary their classroom procedures more frequently from controlled to free types of drills and vice versa are evidently more successful than teachers who stay with the same type of drill for prolonged periods of time.

In view of the positive results obtained in the testing of this hypothesis, we should consider the possibility that the differences in student achievement in these ten classes may have been influenced by the previously discussed factors of sex, college orientation, laboratory practices and student attitude. Information concerning these factors in relation to Variable 10 may be found in Table 11.

TABLE 10

t-Tests for Significance of Difference of Means for Classes Ranked as Top 5 and Bottom 5 on Teacher Behavior Variable 10

(All differences are weighted differences for classes of unequal numbers)

Class # (High)	Mean	n	Class # (Low)	Mean	n	d_i
			Listening test			
8	5.17	12	12	6.30	10	-2.64
15	7.47	17	16	4.34	26	10.06
2	7.24	25	13	5.24	17	6.34
5	5.87	23	6	5.67	25	.65
14	6.62	24	9	6.81	21	-.63
The paired t-test based on $\bar{d}$ yields t = 1.17 with 4 df, ns.						
			Reading test			
8	4.32	12	12	4.37	10	-.11
15	4.02	17	16	2.71	26	4.21
2	3.92	25	13	3.62	17	.95
5	2.38	23	6	3.22	25	-2.91
14	3.80	24	9	3.23	21	1.91
The paired t-test based on $\bar{d}$ yields t = .69 with 4 df, ns.						

TABLE 10 (Continued)

Class # (High)	Mean	n	Class # (Low)	Mean	n	d_i
Grammar test						
8	11.74	12	12	9.84	10	4.44
15	10.76	17	16	7.59	26	10.19
2	10.44	25	13	9.39	17	3.33
5	12.07	23	6	10.33	25	6.04
14	13.48	24	9	8.17	21	17.82
The paired t-test based on $\bar{d}$ yields $t = 3.18$ with 4 df, $p < .05$.						
Writing (Base) test						
8	12.73	12	12	13.07	10	-.79
15	13.05	17	16	7.02	26	19.39
2	18.69	25	13	9.39	17	29.53
5	12.07	23	6	10.33	25	6.04
14	13.48	24	9	8.17	21	17.82
The paired t-test based on $\bar{d}$ yields $t = 2.71$ with 4 df, $p < .05$.						
Writing (Free) test						
8	17.92	12	12	11.00	10	16.17
15	22.94	17	16	3.58	26	62.25
2	47.88	25	13	8.53	17	124.92
5	17.39	23	6	9.40	25	27.74
14	19.13	24	9	8.81	21	34.63
The paired t-test based on $\bar{d}$ yields $t = 2.73$ with 4 df, $p < .05$.						

TABLE 10 (Continued)

Class # (High)	Mean	n	Class # (Low)	Mean	n	d_i
Speaking (Base) test						
8	41.08	6	12	25.04	7	28.95
15	42.77	12	16	23.87	13	47.49
2	46.50	12	13	33.13	10	31.31
5	27.98	9	6	28.10	13	-.28
14	37.50	10	9	35.43	12	4.96
The paired t-test based on $\bar{d}$ yields $t = 2.54$ with 4 df, $p < .05$.						
Speaking (Free) test						
8	3.15	6	12	1.61	7	2.78
15	4.76	12	16	1.64	13	7.84
2	5.40	12	13	1.03	10	10.23
5	4.56	9	6	1.53	13	7.01
14	1.31	10	9	2.19	12	-2.06
The paired t-test based on $\bar{d}$ yields $t = 2.38$ with 4 df, $p < .05$.						

TABLE 11

Variables Possibly Affecting Achievement of Students in Classes Considered in Variable 10

Class #	Proportion of females	College orientation	Lab minutes per week	Attitude pre-test
		Top 5 Classes		
2	.68	.75	30 (out)	77
5	.82	.75	30 (out)	91
8	.83	.80	45 (in)	79
14	.79	.80	45 (in)	76
15	.76	.80	45 (in)	65
		Bottom 5 Classes		
6	.72	.65	45 (out)	80
9	.48	.75	60 (out)	77
12	.50	.35	10 (in)	80
13	.53	.75	40 (in)	74
16	.58	.65	50 (out)	79

As Table 11 indicates, the group of classes of the high-ranking teachers contains, as a whole, a higher proportion of girls than the classes of the low-ranking teachers. However, achievement scores, once adjusted for aptitude differences, showed no significant correlation with proportion of girls to boys. Thus, as has been pointed out already, it was not possible to further adjust the scores on the criterion measures for sex composition. The correlations for the seven criterion variables based only on the ten classes of the teachers considered in Hypothesis 2 (the five highest and five lowest on Variable 10) were:

Listening (Variable 34)	.08
Reading (Variable 35)	-.01
Grammar (Variable 36)	.38
Writing, Base (Variable 37)	.44
Writing, Free (Variable 38)	.39
Speaking, Base (Variable 39)	.38
Speaking, Free (Variable 40)	.48

It will be noted that although these correlations are high, they still fail to reach significance.

Since the data revealed that there was at least a tendency for higher female/male ratios in the classes of the teachers ranking high on Variable 10, the correlation of Variable 10 with the female/male ratios for all 17 teachers taking part in the experiment was examined. It was found that there was, indeed, a significant correlation (.50, $p < .05$). Thus, the female/male ratio, while not significantly correlated with achievement, does correlate significantly with the teaching behavior observed through Variable 10: The higher the number of girls in a class, the more likely the switching from Controlled to Free drill. We can only speculate about the possible implications of this finding. One possible suggestion is the following: The majority of teachers in this study (14 out of 17) are female. Variable 10 is largely an indication of the teacher's willingness to vary classroom behavior in response to student reaction. The higher the female composition of the class, the more likely it becomes that the (female?) teacher is sensitive to student reaction and adjusts the teaching procedures to it.

No clear difference in initial student attitude (Variable 25) emerges from the comparison of the two groups. Three "average" classes (with attitude scores of 77, 79, 76) in the high group are matched by classes with similar scores in the low group (77, 74, 79). The low classes with somewhat higher initial attitude (80, 80) are balanced by one very high score (91) and one very low score (65) in the high group. As will be discussed later in the correlational analysis, initial high attitude scores may have been influenced by certain aspects of teaching behavior. Thus, it was neither possible nor advisable to adjust criterion measures for differences in attitude scores.

The approximate percentage of college oriented students within the schools in which the teachers are operating seems to favor the high group. Both groups contain two classes which came from a 75% college oriented environment. But the three classes from 80% college oriented schools in the high group correspond to classes from only 65%, 65%, and 35% college oriented schools in the low group. Again, it will be recalled that the total correlation of achievement with college orientation did not indicate the necessity for adjusting the criterion measures. Nevertheless, there seems to be a tendency for a higher college orientation of pupils to be found in classes of teachers ranking high in Variable 10. This tendency (a correlation of .35 between Variable 10 and college orientation) does not reach the same significance level as the relation between Variable 10 and the female/male ratio. It does, however, reinforce the suspicion that some of the teaching behaviors that have been observed are influenced by the composition of the classes--a fairly plausible hypothesis if we remember that many behaviors, especially the flexibility measure implied in Variable 10, are really based on interaction rather than an action by the teacher alone. We can also point out that the correlation of Variable 10 with the average language aptitude of the class is .40, approaching the .48 required for a .05 level of significance (.41 is significant at the .10 level only).

Utilization of the language laboratory does not provide any clearcut differentiation between the classes of high- and low-ranking teachers either: Two classes of the high and three of the low group utilize the lab outside of regular classroom activity. While one class in the low group utilizes the lab only infrequently (10 minutes per week) two other classes in this group spend more time in the lab (50 and 60 minutes per week) than any of the high group classes.

We can thus assume that neither college orientation, sex composition, laboratory utilization, nor student attitude could have influenced the difference in achievement between the students of the high and low ranking classes in a very significant way.

Correlational Analysis

The variables with which this study is concerned consist (as has been stated on page 5) of the following:

A. Teacher Behaviors (1-14).
B. Teacher Characteristics (11-24).
C. Student Attitudes (25-32).
D. Criterion Tests of Student Achievement (33-40).

The discussion of the correlations between these variables will follow this plan:

1. Correlations within each group of variables: (a) AA; (b) BB; (c) CC; (d) DD.
2. Correlations between Teacher Behaviors and Teacher Characteristics: AB.
3. Correlation between Teacher Behaviors and Characteristics and Student Attitudes: (a) AC; (b) BC.

4. Correlations between Student Achievement and Teacher Behaviors, Teacher Characteristics, Student Attitudes: (a) AD; (b) BD; (c) CD.

In the discussion of the correlations, we should keep in mind the following significance levels (See Dixon and Massey 1951, p. 460):

	.05	.01
N = 15	.514	.641
N = 16	.497	.623
N = 17	.482	.606

On the correlation matrix tables, correlations which approach significance to the extent that they will be utilized in the interpretation (.1-.05) have been marked [a]; correlations with an .05 significance level are marked *; those with a significance level of .01 and beyond are marked **.

Correlations within each group of variables.

(a) Teacher Behaviors (AA: Appendix IX-a). Variable 1 (Proportion of Time Spent in Drill Activity) is influenced undoubtedly more than any other by the sampling accident. Just how much time within a sixty (4 x 15) minute sample is spent in drill activity may--as such--give comparatively little information about the usual teaching practices. It was, of course, precisely for that reason that most of the variables observed were based on the recording of activities taking place within the drill activity itself. As such, a greater amount of time spent on drill activity would only increase the opportunity to observe certain behaviors. Thus the correlation between Variable 1 and Variable 14 (Variation of Structure) can simply be explained by the fact that a greater amount of time devoted to drill activity increased the opportunity to deal with various structures.

Variable 2 (Average Number of Drills pdm) is a measure of speed and an indication of the number of structures to which the student is exposed. As a measure of speed, it has a built-in relation to Variables 3, 4, 5, 6, and 7 which are also expressed as activities per drill minute. In fact, this relation is expressed in correlations of $(2,3)^a$, $(2,4)^*$, and $(2,7)^a$. That speed (2) should also have a mild correlation with Variation of Structure (14) may again simply be due to an increase in opportunity brought about by speedy presentation. Of more significance is the correlation between Variable 2 and Free Response (8). Evidently, the teacher who allows for free student responses also tends to conduct drill activities in lively ways.

Among the various drill activities measured as ratios over time (3, 4, 5, 6, 7), the first four (Repetition, Substitution, Dialogue and Translation drills) share the common feature that the student is typically not expected to produce any structural change in his response. Linguistically and pedagogically, the techniques are thus--to some extent at least--alternatives of each other. In observations made within a limited time period, we should thus not be surprised to find frequent use of one of these techniques accompanied by comparatively little use of the others. The negative correlations (3,4), $(3,5)^*$, $(3,6)^a$ thus express an expected relationship. The positive correlation of 5 (Dialogue drill) with 13 (Student to Student Interaction) $(5,13)^*$ is also not surprising. Teachers who emphasize the memorization of dialogues are likely to have the dialogues acted out by students, and student interaction--in the context observed in this experiment--consisted largely of dialogue presentation.

Of the time ratios measured, only Variable 7 (Average Conversion Drills pdm) expects a student response structurally different from the stimulus. We can thus expect that Variable 7 should show positive correlations to measures which relate to teaching behaviors that give greater freedom to the student or indicate greater variety of approach. In fact, these relations are indicated by the correlations of 7 with 8 (Free Response)^a;

10 (Proportion of Switched Drill to Exclusive Drill)*; 12 (Use of Visual Aids)[a]; and 14 (Variation of Structures)*. The high correlation of 7 to 9 (Ratio of Free Drills to Controlled Drills)** is, of course, to be expected since the number of Conversion drills make up most of the numerator of the ratio.

Variables 7, 8, 9, 10, 12, and 14 can thus be interpreted as being manifestations of a less directive, more imaginative type of teaching that introduces a somewhat greater choice of expression into the foreign language classroom. In general, the patterns of correlations between these variables support this interpretation: (7,8)[a], (7,9)**, (7,10)*, (7,12)[a], (7,14)*, (8,10)[a], (8,12)**, (8,14)**, (9,10)*, (9,10)*, (9,14)*, (10,12)[a], (10,14)*, (12,14)**. There is little association between any of these variables and other observed behaviors. As already mentioned, (14,1)* and (14,2)[a] may largely express increased opportunity for variations within a fixed time block because of speed of delivery or greater amount of time devoted to drill as such. Variables 7 and 2 have a built-in mathematical relationship since the numerator of 7 is part of the numerator of 2 (7,2)[a]. This leaves (8,2)* as the tie-in between speed (Variable 2, Average Number of Drills pdm) and the complex associated with the "imaginative" or "varied" teaching behaviors.

(b) Teacher Characteristics (BB: Appendix IX-b). Among the measures of teacher characteristics, Variables 15, 16, and 17 (Teacher's Attitude toward the Textbook, Years of Experience in Teaching French, and Number of Months Spent in France) represent data gathered from the questionnaire. The rest of the variables are the scores of the MLA-ETS tests. Here, it should be kept in mind that Variables 18, 19, 20, and 21 represent the results of the Skill tests (Listening, Speaking, Reading, Writing), while Variables 22, 23, and 24 measure the "non-skill" aspects of the teachers' preparation (Applied Linguistics, Civilization and Culture, Professional Preparation).

Probably the most startling aspects of the correlations in the Teacher Characteristics block is that the Teacher's Attitude toward the Textbook (15) correlates negatively with all the scores achieved in the MLA-ETS tests: (15,18), (15,19), (15,20)*, (15,21)*, (15,22)*, (15,23)^{a}, and (15,24)*. The only indication of positive correlation of (15) is with Years of Experience in Teaching French (15,16)^{a}. In other words, the better the teacher performs on the MLA-ETS test, the more critical he tends to be of the textbook! This is not to be construed as a negative criticism of the specific textbook used by the participating teachers, since the teacher who is proficient in the language would be more apt to modify and adapt whatever textbook he is using.

Except for a somewhat less critical attitude toward the textbook, (15,16)^{a}, teaching experience has evidently no relation to any of the other variables measured among the teacher's characteristics. What is perhaps even more surprising is that, except for a mild correlation with the Listening test (17,18), Number of Months Spent in French-Speaking Country has no significant correlation with performance on the Skill tests. One can only speculate about this unexpected finding. Only one teacher (Teacher 2) had spent any appreciable amount of time in France (three years). Five teachers (Teachers 1, 9, 10, 12 and 17) had spent a year in France. None of the teachers was a native speaker of French. Only one of the teachers who had spent one year in France had taken the MLA-ETS test previously while several who had spent less than a year in France had taken the same test several times at NDEA Institutes or in Teacher Training Programs. The scores on the test, therefore, may reflect relative familiarity with the test itself rather than a genuine measure of relative proficiency in French.

The scores on the Skills tests show the high correlations we might expect between performances in Speaking, Listening, Reading and Writing: (18,19)*, (18,20)**, (18,21)**, (19,20)**, (19,21)*, (20,21)**. Similar high correlations between achievements in these tests were found by other investigators (Carroll et al. 1967, p. 52).

Among the non-skill tests, both Applied Linguistics (22) and Civilization and Culture (23) correlate highly with achievement in the skills: (22,18)^a, (22,19)**, (22,20)**, (22,21)**, (23,18)^a, (23,20)**, (23,21)*. Only the Professional Preparation test--the one test which is outside the "French language and culture complex"--shows no significant relation with skills in French. The non-skills test scores do, however, correlate significantly with each other: (22,23)**, (22,24)*, (23,24)*.

(c) Student Attitudes (CC: Appendix IX-c). There are three separate measures involved in the variables concerned with student attitude: The scores obtained in the general attitude questionnaire, students' reactions to a statement concerning intent to continue foreign language study and students' reactions to a statement concerning the importance of Foreign Languages in the curriculum. For each one of these measures there is one score obtained at the beginning of the year, one obtained at the end of the year and one expressing the gain (or loss) from the first to the second administration of the questionnaire. Thus, it should be kept in mind that Variables 25, 28 and 31 refer to initial attitude measures, 26, 29, 32 refer to post measures and 27, 30, 33 refer to differences between initial and post measures.

The correlations (25,28)** and (25,31)** are thus intercorrelations of the initial attitude measures and merely confirm the validity of the instrument as a general predictor of attitudes toward Foreign Language study. Interestingly enough, the correlations between the post measures are less striking than those between the initial measures. General Attitude Post-test (26) still correlates highly with Intent to Continue Post-test (29): (26,29)**. However, it is not significantly correlated with the students' view of the importance of Foreign Languages in the curriculum (Variable 32). Gain in General Attitude (27) is not significantly related to Gain in Intent to Continue (27,30) but does correlate significantly with Gain in Importance of Subject (27,33)*.

High initial scores on all measures tend to correlate negatively with gains, perhaps because the higher the expectation at the beginning of the course, the more likely that subsequent attitude measures will register disappointment rather than fulfilment: (25,27)*, (28,30)*, (31,33)**. The general rank correlations between pre- and post-measures as such, however, remain positive: (25,26)[a], (28,29)*, (31,32)*. Other significant correlations in this block represent, typically, either negative relations between pre-test highs and gains: (25,27)[a], (31,27)*, or positive correlations between post-tests and gains: (26,27)*.

(d) Criterion Measures of Student Achievement (DD: Appendix IX-d). The criterion measures can be divided into those which measure a "basic" proficiency in Listening, Reading, Grammar, Writing, and Speaking (Variables 34, 35, 36, 37, 39) in the sense that only responses learned very directly in the materials of the A-LM program are expected and evaluated in the tests. In addition, there are also the tests of "Free Writing" (Variable 38) and "Free Speaking" (Variable 40) in which the pupil is encouraged to produce supplementary responses.

As can be seen from Appendix IX-d there is generally a very high degree of intercorrelation among all the criterion measures. This is expected in tests which measure different aspects of the knowledge of French and which are all focused on the rather limited amount of French presented in the first semester of the first level. The only two instances in which there is no indication of a positive correlation among the criterion scores are (40,35) and (40,36), the relations of the Free Speaking test (40) with Reading (35) and Grammar (36). The Free Speaking test measures, in addition to control of vocabulary and grammar, the desire, willingness and interest of the student in producing French freely--a factor which is no doubt reflected in this lack of correlation. It will be noted that the correlation between the two "free expression" tests in Speaking and Writing is very high: (38,40)** and that the free expression tests in Speaking and Writing show the expected high correlations with the Base measures in the same skills: (37,38)** and (39,40)*.

Correlations between groups of variables.

(a) Teacher Behaviors and Teacher Characteristics (AB: Appendix IX-e). For reasons already mentioned, it is difficult to give any interpretation to the slight tendency of the Proportion of Time Spent in Drill Activity (1) to correlate positively with the results of the Linguistic tests (1,22)^{a} and negatively with the Teacher's Attitude toward the Textbook: (1,15)^{a}. In general, one would expect speed in conducting drills (2, Average Number of Drills pdm) to have some association with language proficiency. In fact, a significant positive correlation between speed and proficiency is shown only in the Comprehension tests (2,18)*. This again brings up the question to what extent the scores on the proficiency test may be influenced by experience with the tests rather than proficiency in the language. Of other measures which include time ratios and (by implication) speed as one of the factors being measured, Variable 3 (Average Repetition Drills pdm) shows indication of some positive correlation to language skills (3,21)^{a} and to professional preparation (3,24)^{a}. The use of Dialogue drills pdm (a procedure which largely reflects "following the textbook") shows an interesting positive correlation with years of teaching experience (5,16)** and a negative one with language skills (5,21)* and knowledge of linguistics (5,22)*.

The use of Free Response (8) shows a very high positive correlation with amount of time spent in France (8,17)**. The measures of variation of drills (10), on the other hand, shows a tendency to correlate negatively with years of teaching experience (10,16)[a] as well as with attitude toward the textbook (10,15)*.

One aspect of classroom behavior which was included in the observations to identify the less imaginative and less independent teacher, namely, Reference to Book (11), shows a rather disquieting tendency to correlate positively with teaching experience (11,16)[a] and an expected tendency to correlate negatively with the amount of time spent in France (11,17) and various language skills (11,18)*, (11,19)[a], (11,20)[a], (11,21)[a].

The Use of Visual Aids (12), on the other hand, does tend to correlate positively with time spent abroad (12,17)[a] and with language proficiency (12,18)*, (12,21)[a].

The use of student interaction (13) which, as we have explained before, has some association with the use of dialogue drill (5) follows in general the pattern of the frequency of dialogue drill: a tendency to correlate positively with teaching experience (13,16)[a] and to correlate negatively with performance in the MLA-ETS tests, especially the non-skill tests: (13,22)*, (13,23)[a], (13,24)[a]. Use of variation of structure (14) shows no significant relation to any of the teacher characteristics with the possible exception of a mild positive correlation with time spent in France (14,17)[a].

The consideration of the relation of classroom behaviors to teacher characteristics may thus be summarized as follows: The behaviors which we can consider as indicative of more "independent," less constrained kind of classroom behavior (7, 8, 9, 10, 12, 14) show some tendency to correlate positively with time spent in France: (8,17)**, (12,17)[a], (14,17)[a] and to some extent with language skills (8,18)[a], (12,18)*. Behaviors which can be associated with a less imaginative, more "bookbound" type of teaching on the other hand (e.g., 5, Average Dialogue Drill pdm and 11, Reference to Book) tend to show a positive relation to teaching experience (5,16)**, (11,16)[a], and a negative relation to performance on the MLA-ETS tests (5,21)*, (5,22)*, (11,18)*, $(11,19)^a$, $(11,20)^a$, $(11,21)^a$ as well as with time spent in France (11,17)*. What emerges is thus, on the one hand, the picture of a teacher who is less bookbound, uses more variety of drills, is more critical of his text, and has spent more time in France. His counterpart is a teacher who is less critical of the text, has spent less time in France, is less imaginative and who provides less variety in the instructional process. For better or worse, the latter type of teacher tends to have more teaching experience than the former.

(b) Teacher Behaviors and Student Attitudes (AC: Appendix IX-f). The influence of teaching behavior on attitude measures, especially initial attitude measures, is difficult to assess. Since the initial attitude measures were obtained during the first week of exposure to French they probably reflect attitudes formed before the pupil entered the French class. However, the frequency of the use of Repetition drills does show negative relations with two attitude pre-measures: (3,25)* and (3,31)*. This is perhaps a result of overexposure to repetition in the first week of the course. On the other hand, both pre- and post-attitude measures correlate positively with the use of Translation drills $(6,25)^a$, (6,26)*, (6,31)*, and (6,32)*. The use of English for the purpose of eliciting French responses in drills is undoubtedly associated with a parallel use of English in providing equivalents and translations for French structures in the dialogues and the drills. We can infer from these correlations that there may perhaps be some motivational advantages in providing English equivalents, at least in the early stages when students are ill at ease in the language.

It is of interest to note that the qualities associated with more inventive classroom behavior do not necessarily correlate positively with attitude measures. As a matter of fact, Variable 9 (Ratio of Free Drills to Controlled Drills) tends to correlate negatively: $(9,29)^a$, $(9,33)^a$, as does Variation of

Structure: (14,30)[a]. However, the most significant relation of all classroom behaviors with student attitude is the negative correlation shown by Reference to Book (11) and General Attitude Post-test (26) as well as to the gains made in the post-test: (11,26)**, (11,27)**. Of all the correlations between classroom behaviors and attitude measures, the highest is the negative one between use of the open book and gains in the overall attitude score! This correlation can be quite easily understood if we remember that the reference to the open book can be interpreted as symptomatic of the less imaginative, "bookbound" approach to language teaching.

(c) Teacher Characteristics and Student Attitudes (BC: Appendix IX-g). Of the teacher characteristics, neither experience nor attitude toward the text nor foreign residence show any relation to the attitude measures. Keeping in mind the caution against interpreting pre-attitude measures, and placing most confidence in post measures and especially relative gain scores, we can easily see that the most significant relation between attitudes and teacher characteristics is described by the positive correlation between gains in attitude and the teacher's knowledge of the foreign language: (27,18)**, (27,19)*, (27,20)**, (27,21)**. Interestingly, the gain in the score of intent to continue (30) does not show similar correlations, while gains in scores on the assessment of the importance of Foreign Languages in the curriculum (33) do: (33,18)[a], (33,19)[a] (33,20)*, (33,21)[a]. Variable (33) shows, also, positive correlations with the teacher's knowledge of Applied Linguistics and Culture and Civilization: (33,22)[a], (33,23)*. The one test score which has no relation to attitude measures is furnished by the Professional Preparation test.

Thus, the striking results of the analysis of teacher behaviors and characteristics and student attitudes can be summarized quite succinctly: The most favorable attitudes and attitude changes toward foreign language study are produced by teachers who know the language; the most negative results are produced by teachers who teach from open books!

(d) Teaching Behaviors and Student Achievement (AD: Appendix IX-h). Neither the amount of time spent in drill nor the number of Substitution or Repetition drills pdm show a significant relationship to criterion measures of Student Achievement. There is a mildly positive correlation of speed (Average Number of Drills pdm) with one of the criterion measures, namely, writing: (2,38)^{a}. In general, the frequency of drill activities not involving changes in structure show either no significant correlation or negative correlations with achievement measures, e.g., (5,36)^{a}, (5,37)*, and (6,39)*. Other measures correlating negatively with achievement are Student to Student Interaction: (13,35)*, (13,36)^{a} (largely associated with the use of dialogue drills) and Reference to Book: (11,34)^{a}, (11,37), and (11,40)**.

The Teacher Behaviors showing positive correlations with criterion measures are without exception the ones that we have characterized as being indicative of the more flexible, imaginative teacher and/or involving greater freedom in the manipulation of constructions on the part of the student.

Of all the Teaching Behaviors, the highest number of significant correlations to achievement measures is shown by the Proportion of Switched Drill to Exclusive Drill and the number of Free Responses: (10,36)*, (10,37)^{a}, (10,38)*, (10,39)**, (10,40)**, (8,37)**, (8,38)**, (8,39)*, (8,40)**. It should be noted that the correlations include not only the Free Speaking and Free Writing measures, but also the "basic" measures of Speaking (39), Writing (37) and Grammar (36). Other positive correlations with achievement measures are shown by Conversion drills pdm (7), Ratio of Free to Controlled drill (9), Use of Visual Aids (12) and Variation of Structure (14): (7,37)^{a}, (7,38)^{a}, (7,39)^{a}, (9,39)^{a}, (12,38)^{a}, (12,39)*, (12,40)**, (14,36)^{a}, and (14,37)^{a}.

TABLE 12

Correlations of Criterion Measures of Student Achievement with Teacher Behaviors

		(34) Listening	(35) Reading	(36) Grammar	(37) Writing, Base	(38) Writing, Free	(39) Speaking, Base	(40) Speaking, Free
Proportion of Drill Time	(1)							
Average Drills per minute	(2)					a		
Repetition drills pdm	(3)							
Substitution drills pdm	(4)							
Dialogue drills pdm	(5)			-a	-*			
Translation drills pdm	(6)						-*	
Conversion drills pdm	(7)				a	a	a	
Free responses pdm	(8)				**	**	*	**
Ratio of Control/Free drill	(9)						*	
Ratio of Switched to Excl.	(10)			**	a	*	**	**
Reference to Book	(11)	-a			-a			-**
Use of Visual Aids	(12)					a	*	*
Student/student interaction	(13)		-*	-a				
Variation of Structures	(14)				a	a		

a Approaching significance
* Significant at .05 level
** Significant at .01 level

The correlation analysis thus tends to confirm results shown in the investigation and discussion of the main hypotheses. Flexibility on the part of the teacher improves not only the scores on the Free Speaking and Free Writing criteria but also correlates positively with achievement on other criterion measures. An emphasis on the dialogue drills and translation drills tends to have a depressing effect on achievement on all measures. In general, the characteristics associated with the more "liberal," more imaginative, less directive teacher are associated with pupil achievement, with the highest correlations produced by the use of free response, frequency of variation from one type of drill to another and the use of visual aids.

(e) Teacher Characteristics and Student Achievement (BD: Appendix IX-i). Among Teacher Characteristics, there is only one which has a significant negative relationship to criterion measures of Student Achievement, i.e., Years of Experience in teaching French (16) correlates negatively with achievement in the Grammar and Base Writing tests: (16,36)*, (16,37)*. (See Table 14.) The significant positive correlations with achievement are shown by length of residence in France (17) and teacher's proficiency on the MLA-ETS comprehension test (18): (17,34)^a, (17,37)*, (17,38)**, (17,39)^a, (17,40)^a, (18,34)^a, (18,37)^a, (18,38)*, (18,39)*, (18,40)**. Both of these variables (17 and 18) can of course be interpreted as indications of language proficiency. It is interesting to note, however, that the other language proficiency tests of the MLA-ETS battery do not show similar high correlations with Student Achievement. The Reading test (20) and Writing test (21) show some almost significant correlations (20,40)^a, (21,39)^a; the Speaking test scores show no significant correlations with Student Achievement whatsoever. The latter fact is especially surprising in view of the high correlations shown by the other audio-lingual skill: Listening. Two possible explanations suggest themselves: (a) as mentioned previously, results in the Speaking test may reflect familiarity with the test, since the scores are based on

TABLE 13

Correlations of Criterion Measures of Student Achievement and Teacher Characteristics

		(34) Listening	(35) Reading	(36) Grammar	(37) Writing, Base	(38) Writing, Free	(39) Speaking, Base	(40) Speaking, Free
Attitude toward Textbook	(15)							
Teaching Experience	(16)			-*	-*			
Time spent in France	(17)	a			*	**	a	a
MLA-ETS Listening Test	(18)	a			a	*	*	**
MLA-ETS Speaking Test	(19)							
MLA-ETS Reading Test	(20)							a
MLA-ETS Writing Test	(21)						a	
MLA-ETS Linguistics	(22)							
MLA-ETS Culture & Civil'n	(23)							
MLA-ETS Professional Prep.	(24)							

a Approaching significance
* Significant at the .05 level
** Significant at the .01 level

responses to relatively few (usually pictorial) stimuli which can be remembered relatively easily by the test-taker. Those teachers in the experiment who had been participants at a second level NDEA Institute in France, for example, would have taken the MLA-ETS test four times and would remember the pictorial stimuli sufficiently to score highly on that part of the battery. (b) Achievement on the Speaking test--based on a performance in front of a microphone--may reflect, in addition to a knowledge of French, various personality characteristics connected with the ability to perform in this particular type of testing situation. No relation to Student Achievement is shown by the non-skill tests of the MLA-ETS battery (22, 23, 24). Thus, with the possible exception of the relationship between the Listening test and Student Achievement, the results of this study are in substantial agreement with the findings of Smith and Berger (1968, p. 133) who also found little correlation between pupil achievement and the teachers' performance on the MLA Proficiency Tests.

(f) Student Attitudes and Criterion Measures of Student Achievement (CD: Appendix IX-j). From Table 14, we note a tendency for initial attitude (25, 28, 31) to correlate negatively with criterion measures: (25,35)*, (25,36)*, (25,39)*, (28,34)*, (31,35)*, (31,36)[a], (31,39)[a]. However, the post-measure of at least one attitude variable (assessment of the importance of Foreign Languages in the curriculum) also shows a negative correlation with Student Achievement: (32,35)**, (32,36)**, (32,37)*, (32,39)*. Obviously, the pupil's assessment of the importance of Foreign Languages in the curriculum is not positively related to achievement at the level of first semester A-LM, nor do the scores indicating Intent to Continue (28, 29, 30) reflect any relation to achievement.

Perhaps the most valid attitude test scores are those furnished by the General Attitude Post-test (26) and Gain scores (27) of the total attitude scale, which do show positive

TABLE 14

Correlations of Criterion Measures of Student Achievement and Student Attitudes

		(34) Listening	(35) Reading	(36) Grammar	(37) Writing, Base	(38) Writing, Free	(39) Speaking, Base	(40) Speaking, Free
General Attitude pre-test	(25)		-*	-*			-*	
General Attitude post-test	(26)							*
Gains from pre- to post-test	(27)							*
Intent to continue (pre)	(28)	-a						
Intent to continue (post)	(29)							
Gain from pre- to post-test	(30)	a						
Importance of subject (pre)	(31)		-*	-a				
Importance of subject (post)	(32)		-**	-**	-*		-*	
Gains from pre- to post-test	(33)							

a Approaching significance
* Significant at .05 level
** Significant at .01 level

correlation with achievement. However, significance is reached only in their correlation with the Free Speaking test: (26,40)*, (27,40)*. This result is not surprising. The Free Speaking test scores reflect more than any others a positive desire and interest in speaking the language and measures, undoubtedly, a motivational factor in addition to language skills.

Summary of correlational analysis.

The relation of all the variables measured--especially Teacher Behaviors and Characteristics--to Student Achievement may be summarized as follows: The composite picture of the "successful teacher" as it emerges from this study is that of a person who:

(1) Gives students the opportunity to respond freely.
(2) Switches frequently from one kind of drill to another, particularly to those requiring linguistic manipulation.
(3) Uses visual aids.
(4) Has spent time in France.
(5) Performs well on the listening part of the MLA-ETS tests.

Furthermore, the successful teacher identified by this analysis is a person who *does not*:

(1) Overemphasize the use of Dialogue drills (student interaction drills with fixed responses).
(2) Refer frequently to his open book during drill activities.
(3) Have a great deal of teaching experience (!).

Some teaching behaviors and characteristics found to be irrelevant to pupil success include:

(1) Speed of drilling (number of stimuli to which the pupil is exposed).
(2) Insistence on a high number of Repetition and Substitution drills.
(3) Teacher's attitude toward the textbook.
(4) Achievement in the non-skills tests of the MLA-ETS battery.

Summary and Implications

Development

In interpreting the results of this study for further development in the training of foreign language teachers and the construction of language learning curricula, it must first of all be kept in mind that the curriculum itself was not one of the variables under consideration, but a constant. Thus, if certain variables (e.g., the frequency of the use of Repetition drills or Substitution drills) turned out not to be correlated to pupil success and others (e.g., the use of Free Response and Visual Aids) showed high correlation to pupil success, we cannot conclude that the latter are "good" teaching procedures while the former are "bad." What must be considered in assessing the correlation of student achievement with teaching behaviors is that certain behaviors are the unavoidable part and parcel of the A-LM curriculum itself. It is the emphasis on these behaviors (e.g., Substitution drill, Repetition drill, Dialogue drill) which seems to have no bearing on the relative achievement of the pupil. The behaviors which make the difference are the ones which are not definitely and unavoidably dictated by the curriculum itself: Free Response drills, visual aids, flexibility in shifting from one type of drill to another. Again, there is no evidence in the results of this study to indicate that one should do nothing but use visual aids or engage in Free Response drill and should not use Repetition drills or Substitution drills. But the better results were obtained by the pupils of those teachers who went beyond the procedures strictly prescribed by the curriculum, teachers who were concerned with supplementing the curriculum rather than merely implementing it.

As far as the selection and training of foreign language teachers are concerned, the results of this study point in these directions:

(1) A certain propensity for innovation and flexibility is evidently a desirable characteristic in the foreign language teacher (and probably in _any_ teacher) and as far as possible individuals with this characteristic should be selected and attracted to the profession.

(2) Residence in the foreign country provides a highly desirable background. To what extent residence in the foreign country produces the confidence and skill of the successful teacher or to what extent it is in itself the result of characteristics which make a good language teacher is difficult to determine.

(3) Training procedures should concentrate not only on implementing methods, but also on the skills of supplementing them. The findings of this study suggest the hypothesis that the efficiency of the individual teacher increases with the amount of his _personal_ stake and _personal_ contribution to the instructional processes. With the increasing use of language laboratories, teaching machines, and perhaps computer based instruction it will also become increasingly easier to completely control and standardize large parts of the curriculum and the teaching procedures used in it. The question of supplementing rather than implementing is thus likely to attain greater importance and may indeed, in the not too distant future, become the major question to which the teacher has to address himself.

Research

The need for replication of the present study must be stressed especially in connection with the problems of sampling teaching behavior which were inherent in this study. Replication in other languages and at different levels are needed in order to confirm the concept of the superiority of the innovative, flexible, book- and curriculum-free teacher that has emerged from this investigation.

Other research suggested by this investigation lies in reconsideration of the available data and observations. What we might call the "macro-model" of research in teaching has, indeed, proved fairly ineffective in the past and there is undoubtedly a great deal of merit in replacing this macro-model with a micro-paradigm. However, it seems to us that the technology available to us now--especially the preservation of observations on video-tape--may well infuse new life into the macro-model just as it has given impetus and inspiration to the micro-model. The correlational analysis and hypotheses on which this study is based have allowed us in turn to advance the hypothesis that the successful language teacher is the non-curriculum (or book) bound, flexible, innovative person. This hypothesis is not completely identical with the research hypotheses of the study, but the fact that the observations on which this study is based are still available may enable us to refine our observational instrument in terms of new hypotheses. It is undeniable that some of the observational variables used in our study were in fact not good indicators of the independent, innovative behavior of the teacher. Conversion drill, for example, while linguistically quite different from Repetition and Dialogue drill, is rather strictly prescribed by the method used in the curriculum and does not require the generation of new, non-predictable utterances by the student. The use of the Conversion drill versus the Substitution drill is thus not an index of innovative, independent behavior on the part of the teacher. Yet Substitution and Repetition drill on the one hand and Conversion drill on the other were used in this study to draw conclusions as to the frequency of varied teaching behaviors. If we assume that innovation, independence, and flexibility are characteristics of the good teacher, then we can try to evaluate the available samples of teaching behaviors according to categories which relate to these characteristics even more directly than the ones used in this study. Among the categories we might suggest at this point are included those of any drills or structures not directly associated with the curriculum (or textbook) and the frequency in adjustments in teaching procedure in response to student reaction. The recent development by Dr. Gertrude Moskowitz of Flanders' Interaction Analysis system for Foreign Language Teachers (Moskowitz 1968), for instance, can provide a sensitive instrument for the observation and rating of just this kind of teacher behavior. Such a process of the generation of new hypotheses and successive refinements of instruments based on already available data and hypotheses may very well be one of the more promising avenues of approach in our research.

The ultimate value of educational research lies not in finding definitive answers to hypothetical questions nor in defending certain assumptions, but in the generation and testing of new and meaningful hypotheses. The conservation of classroom observations made possible by the videotape recorder opens broad and exciting avenues of approach that can lead us to a greater understanding of the educational process.

REFERENCES

A-LM French, Level One. New York: Harcourt, Brace & World, Inc., 1961.

Brisley, Leonard, Dallaccio, Carl, Funke, Francis J., Hamlin, Donald J. and Leamon, M. Phillip. Good Teaching Practices: A Survey of High School Foreign Language Classes. Reprinted from Reports of Surveys and Studies in the Teaching of Modern Foreign Languages. The Modern Language Association, 1961.

Carroll, John B. and Sapon, S. M. Modern Language Aptitude Test. New York: The Psychological Association, 1959.

Carroll, John B., Clark, John L. D., Edwards, Thomas M. and Handrick, Fannie A. The Foreign Language Attainments of Language Majors in the Senior Year. Final Report for Contract OE 4-14-048, U. S. Office of Education, Department of Health, Education and Welfare, 1967.

Cronbach, L. J. Coefficient Alpha and the Internal Structure of Tests. Psychometrika, vol. 16, pp. 297-334, 1951.

Dixon, W. J. and Massey, F. J., Jr. Introduction to Statistical Analysis. New York: McGraw-Hill Book Co., 2nd Ed., 1957.

Dixon, W. J. (Ed.) Biomedical Computer Programs. Berkeley and Los Angeles: University of California Press, 1967.

Gage, N. L. Paradigms for Research on Teaching in Gage, N. L. (Ed.), Handbook of Research on Teaching. Chicago: Rand, McNally & Co., 1963.

Gage, N. L. An Analytical Approach to Research on Instructional Methods. Phi Delta Kappan, vol. 49, pp. 601-606, June, 1968.

Gage, N. L., Belgard, Maria, Dell, D., Hiller, J. E., Rosenshine, B. and Unruh, W. R. Explorations of the Teacher's Effectiveness in Explaining. Technical Report No. 4, Stanford Center for Research and Development in Teaching, Stanford University, 1968.

Hayes, Alfred S., Lambert, Wallace E. and Tucker, G. Richard. Evaluation of Foreign Language Teaching. Foreign Language Annals, vol. 1, pp. 22-44, October, 1967.

Kuder, G. F. and Richardson, M. W. The Theory of the Estimation of Test Reliability. Psychometrika, vol. 2, pp. 151-160, 1937.

Mathieu, Gustav. On the Anatomy of Listening-Comprehension. The French Review, vol. 35, pp. 50-53, October, 1961.

Modern Language Association. Foreign Language Proficiency Tests for Teachers and Advanced Students. Princeton, New Jersey: Educational Testing Service, 1966.

Moskowitz, Gertrude. The Effects of Training Foreign Language Teachers in Interaction Analysis. Foreign Language Annals ACTFL, vol. 1, pp. 218-235, March, 1968.

Politzer, Robert L. Practice Centered Teacher Training: French. Technical Report No. 1, Stanford Center for Research and Development in Teaching, Stanford, California, 1966.

Rivers, Wilga M. The Psychologist and the Foreign Language Teacher. Chicago: The University of Chicago Press, 1964.

Scherer, George A. C. and Wertheimer, Michael. A Psycholinguistic Experiment in Foreign Language Teaching. New York: McGraw-Hill & Co., 1964.

Smith, Philip D. and Baranyi, Helmut A. A Comparison Study of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment. Final Report Project #7-0133, U. S. Office of Education, Department of Health, Education, and Welfare, October 1968.

Smith, Philip D. and Berger, Emanuel. An Assessment of Three Foreign Language Teaching Strategies Utilizing Three Language Laboratory Systems. Final Report, Project 5-0683, United States Office of Education, Department of Health, Education, and Welfare, January, 1968.

Stanford Center for Research and Development in Teaching. First Annual Report, April, 1967.

Valette, Rebecca M. Modern Language Testing. New York: Harcourt, Brace & World, Inc., 1967.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co., 1962.

APPENDIX I

One-Way Analysis of Variance for Rater Reliability

Measures of Frequency of Controlled Drills per Drill Minute for Four Teachers as Rated by Three Observers

Source	df	SS	ms
Between people	3	27.61	9.20
Within people	8	1.50	.18
Total	11	29.11	
Reliability: .99*			

Measures of Minutes of Exclusive Drill for Four Teachers as Rated by Three Observers

Source	df	SS	ms
Between people	3	57.66	19.22
Within people	8	5.34	.66
Total	11	63.00	
Reliability: .97*			

*Formula for Reliability: $1 - \frac{\text{ms Within people}}{\text{ms Between people}}$

See Winer (1962) p. 128.

APPENDIX II

VIDEOTAPE OBSERVATION RATING SHEET

Teacher # ____________
Tape # ____________
Rater ____________

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Repetition															
2. Substitution															
3. Conversion															
4. Free Response															
5. Dialogue drill															
6. Translation															

GENERAL IMPRESSION

	Never	Occas.	Frequ.
Direct reference to book			
Use of visual aids			
Student/student interaction			
Variation of structures			

COMMENTS:

APPENDIX III

Ratings of Teacher Behaviors: Block A

Teacher	Variable													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.42	4.60	3.0	.52	0	.12	.98	0	.26	.19	0	4	2	4
2	.80	10.02	4.71	1.50	0	0	2.89	.91	.61	.96	2	6	2.7	6
3	.73	9.16	6.47	1.59	.09	.32	.68	0	.08	.24	5.3	2.7	2	3.3
4	.58	4.57	2.85	0	.37	.08	1.26	0	.38	.24	5.5	4	2	3.5
5	.85	6.32	1.37	1.76	0	1.81	1.16	.21	.28	.73	2	4.7	3.3	4.7
6	.93	5.13	2.39	2.48	0	0	.25	0	.05	.10	4.7	3.5	2	4
7	.77	4.20	3.45	.43	.17	0	2.30	0	.56	.24	4.5	3.5	3	4
8	.97	7.23	4.04	.66	0	.20	2.05	.27	.47	1.59	4	4.5	2	6
9	.62	6.21	1.29	1.12	1.86	0	1.43	0	.30	.16	4	4	4.70	4
10	.93	5.26	3.07	.37	0	.34	1.48	0	.39	.56	2.5	4.7	3	4
11	.68	8.17	3.02	3.19	0	.61	1.34	0	.19	.17	5	4	2	4.5
12	.68	6.22	2.56	.85	.34	1.63	.80	0	.15	.11	3	2	2	3.5
13	.75	6.35	4.62	.53	0	.15	1.04	0	.20	.13	4	4.5	2.5	3.5
14	.85	8.09	2.37	2.37	0	.02	3.86	.02	.92	.89	2.5	3.5	3	4.5
15	.55	7.78	6.47	.61	0	.24	1.57	0	.25	1.54	2	4.5	2	3
16	.90	7.22	.81	2.09	1.62	1.42	1.24	.02	.21	.02	4.5	3.5	2.5	4
17	.93	9.66	4.19	2.14	.38	.34	2.37	.23	.37	.37	2.5	5.5	3.5	5.5

APPENDIX IV

Class Means, Standard Deviations, Maximum, Minimum

MLAT Scores for Samples Used in
Computing Variables 34-38

Class	n	Class mean	sd	Maximum	Minimum	Range
1	14	94.07	35.02	168	48	120
2	25	102.28	18.08	139	65	74
3	17	82.88	24.04	126	44	82
4	19	93.42	17.02	141	65	76
5	23	93.52	21.20	135	59	76
6	25	97.28	23.59	149	61	88
7	16	79.00	28.35	137	30	107
8	12	84.67	25.98	138	48	90
9	21	99.62	21.38	143	53	90
10	12	94.25	25.72	137	49	88
11	26	97.65	19.23	132	60	72
12	10	88.70	29.44	127	38	89
13	17	87.47	26.39	128	46	82
14	24	92.79	16.99	128	59	69
15	17	115.00	15.17	138	88	50
16	26	82.65	19.27	137	43	94
17	16	83.94	22.39	132	49	83

MLAT Scores for Samples Used in Computing Variables 39-40

Class	n	Class mean	sd	Maximum	Minimum	Range
1	11	100.18	37.28	168	48	120
2	12	109.42	19.93	139	69	70
3	5	93.00	33.76	126	44	82
4	10	94.40	19.74	141	65	76
5	9	89.78	21.41	121	60	61
6	13	100.69	26.03	149	61	88
7	5	82.40	22.32	110	57	53
8	6	92.83	28.41	138	66	72
9	12	105.25	16.23	134	84	50
10	8	92.13	27.65	137	49	88
11	12	103.25	19.50	132	73	59
12	7	96.43	32.21	127	38	89
13	10	96.80	20.64	128	73	55
14	10	90.70	15.94	116	59	57
15	12	119.25	15.96	138	88	50
16	13	91.85	19.77	137	64	73
17	12	83.58	25.58	132	49	83

APPENDIX V

Adjusted Class Means on Criterion Tests in Rank Order

(Figures at left of each column signify class number)

Listening		Reading		Grammar		Writing, Base		Writing, Free		Speaking, Base		Speaking, Free	
11	(7.96)	11	(4.91)	8	(11.74)	2	(18.69)	2	(47.88)	2	(46.50)	2	(5.40)
15	(7.47)	1	(4.45)	1	(11.11)	11	(16.33)	11	(27.81)	15	(42.77)	15	(4.76)
2	(7.24)	12	(4.37)	11	(10.89)	14	(13.98)	15	(22.94)	8	(41.06)	5	(4.56)
10	(6.92)	8	(4.32)	15	(10.76)	1	(13.38)	14	(19.13)	11	(40.43)	8	(3.15)
9	(6.81)	15	(4.02)	14	(10.68)	12	(13.07)	8	(17.92)	14	(37.55)	10	(2.92)
1	(6.78)	2	(3.92)	2	(10.44)	15	(13.05)	5	(17.39)	1	(36.47)	1	(2.57)
14	(6.62	14	(3.80)	4	(10.27)	8	(12.73)	10	(14.67)	4	(35.53)	9	(2.19)
7	(6.31)	10	(3.69)	12	(9.54)	5	(12.07)	1	(13.86)	9	(35.43)	11	(1.83)
12	(6.30)	13	(3.62)	13	(9.20)	17	(11.07)	12	(11.00)	13	(33.13)	16	(1.64)
5	(5.87)	3	(3.40)	10	(9.14)	4	(10.97)	6	(9.40)	10	(31.03)	12	(1.61)
6	(5.68)	4	(3.28)	6	(8.37)	7	(10.72)	9	(8.81)	17	(30.11)	6	(1.53)
17	(5.31)	9	(3.23)	9	(8.22)	10	(10.70)	3	(8.58)	6	(28.10)	14	(1.31)
13	(5.24)	6	(3.22	7	(8.05)	6	(10.33)	13	(8.53)	5	(27.98)	17	(1.09)
18	(5.17)	17	(3.06)	17	(8.03)	13	(9.39)	4	(7.26)	7	(25.63)	13	(1.03)
4	(5.05)	7	(2.96)	3	(7.62)	3	(9.07)	7	(7.13)	12	(25.04)	7	(.80)
16	(4.34)	16	(2.71)	5	(7.60)	9	(8.17)	16	(3.58)	16	(23.87)	4	(.60)
3	(4.00)	5	(2.38)	16	(7.59)	16	(7.02)	17	(3.44)	3	(25.58)	3	(.14)

APPENDIX VI

Teacher Questionnaire

STANFORD CENTER FOR RESEARCH AND DEVELOPMENT IN TEACHING

Name ______________________ Sex F ___ M ___

School ______________________ Age ___

School District ______________________

City ______________________

1. In what State did you spend most of your childhood?

____

2. Was your mother a native speaker of a foreign language?

Yes ______ No ______

If yes, what foreign language? ______________________

3. Was your father a native speaker of a foreign language?

Yes ______ No ______

If yes, what foreign language? ______________________

4. Is your spouse a native speaker of a foreign language?

Yes ______ No ______

If yes, what foreign language? ______________________

5. Have you ever spent time abroad in a predominantly French country? Yes ________ No ________

If yes, how much time?

Less than a month ____ From one to three months ____

Six months ____ A year ____ Over a year ____

6. Where did you get your B.A.? ________________ When ______ Year

__________________ Major Subject __________________ Minor Subject

7. Where did you get your M.A.? ________________ When ______ Year

__________________ Major Subject __________________ Minor Subject

8. Have you attended an NDEA F. L. Institute? Yes ___ No ___

Where ____________________________ When ______ Year

9. Have you attended any other institutes or language workshops?

Yes ___ No ___

Where ____________________________ When ______ Year

10. How many years of teaching experience have you had?

Teaching (Total) ________________ year(s)

Teaching F. L. ________________ year(s)

Teaching French ________________ year(s)

Teaching other (Name subject) ____________________

11. Please indicate your impressions of A-LM as a textbook. Place a check mark in the space closest to the description you agree with. If you undecided or if your impression is neutral, place a check mark in the center space.

effective	___	___	___	ineffective
productive	___	___	___	wasteful
imaginative	___	___	___	unimaginative
rich in content	___	___	___	poor in content
challenging	___	___	___	unchallenging
convenient	___	___	___	inconvenient
adaptable	___	___	___	not adaptable
interesting	___	___	___	boring
pleasing	___	___	___	irritating
flexible	___	___	___	rigid

APPENDIX VII

Mary Dufort Student Attitude Scale

STANFORD CENTER FOR RESEARCH AND DEVELOPMENT IN TEACHING

FOREIGN LANGUAGE ATTITUDE SCALE

(Adapted from Foreign Language Scale, ©1962 Mary Dufort)

This is a questionnaire which will give you an opportunity to express how you feel about the foreign language you are studying. The responses you give to these statements are important for research purposes. YOUR RESPONSES WILL NOT BE SEEN OR EVALUATED BY YOUR CLASSROOM TEACHER OR BY ANY OTHER OFFICIAL IN YOUR SCHOOL.

Please do not write in this booklet but use the answer sheet provided. Read each statement carefully and put down your first reaction by writing a number from 1 to 4 in the space provided. For example, a statement might read:

Sample: Football is an important school activity.
If you do not agree at all, write 1
If you agree a little bit, write 2
If you agree quite a bit, write 3
If you agree very much, write 4

Choose only one response for each statement, but be sure to respond to every statement with a number from 1 to 4. Do not use 0 .

REMEMBER: YOUR RESPONSES WILL HAVE NO INFLUENCE WHATEVER ON YOUR CLASS GRADE.

1. I like studying French.
2. I would like to learn more than one foreign language.
3. I like to practice French on my own.
4. Most people enjoy learning a foreign language.
5. Everyone in school should take a foreign language.
6. French is interesting.
7. It is too bad that so few Americans can speak French.
8. Anyone who can learn English can learn French.
9. I would like to travel in a country where French is spoken.

10. The way French people express themselves is very interesting.

11. French is an easy language to learn.

12. I would like to be a French teacher.

13. I would like to take French again next year.

14. The French I am learning will be useful to me.

15. I would like to know French-speaking people of my own age.

16. Students who live in French-speaking countries are just like me.

17. I'm glad French is taught in this school.

18. My parents are pleased that I'm learning French.

19. I like to hear French people talk.

20. French is one of my most interesting subjects.

21. Studying French helps me to understand people of other countries.

22. I think everyone in school should study a foreign language.

23. Americans really need to learn a foreign language.

24. What I learn in French helps me in other subjects.

25. Learning French takes no more time than learning any other subject.

26. Sometimes I find that I'm thinking in French.

27. My friends seem to like taking French.

28. I'm glad that I have the opportunity to study French.

29. I use French outside the classroom.

30. I'm looking forward to reading French books on my own.

31. I would like to study more French during the next school year.

32. French is one of the most important subjects in the school curriculum.

Name of Student: ____________________ Class: ____________

School: ____________

Answer Sheet

Foreign Language Attitude Scale

1. 1 2 3 4
2. 1 2 3 4
3. 1 2 3 4
4. 1 2 3 4
5. 1 2 3 4
6. 1 2 3 4
7. 1 2 3 4
8. 1 2 3 4
9. 1 2 3 4
10. 1 2 3 4
11. 1 2 3 4
12. 1 2 3 4
13. 1 2 3 4
14. 1 2 3 4
15. 1 2 3 4
16. 1 2 3 4
17. 1 2 3 4
18. 1 2 3 4
19. 1 2 3 4
20. 1 2 3 4
21. 1 2 3 4
22. 1 2 3 4
23. 1 2 3 4
24. 1 2 3 4
25. 1 2 3 4
26. 1 2 3 4
27. 1 2 3 4
28. 1 2 3 4
29. 1 2 3 4
30. 1 2 3 4
31. 1 2 3 4
32. 1 2 3 4

APPENDIX VIII

Criterion Tests of Student Achievement

STANFORD UNIVERSITY
SCHOOL OF EDUCATION

Student ____________________

Teacher ____________ School ____________

LISTENING TEST

You will hear a series of ten questions. After each question you will hear three responses, only one of which is correct. Decide which response is the correct one and fill in the 0 in the column corresponding to its number.

Example: Qui est-ce?

1. C'est tout droit.
2. C'est un ami.
3. C'est aujourd'hui mercredi.

Since C'est un ami is the only possible correct response, you would fill in the 0 under 2. You will hear the questions and responses only once, so listen carefully.

	1	2	3
1.	0	0	0
2.	0	0	0
3.	0	0	0
4.	0	0	0
5.	0	0	0
6.	0	0	0
7.	0	0	0
8.	0	0	0
9.	0	0	0
10.	0	0	0

STANFORD UNIVERSITY
SCHOOL OF EDUCATION

LISTENING TEST (Tape script)

1. Combien font deux et trois?

    1. Bien, merci. Et toi?
    2. Oui, ils sont là-bas.
    3. Deux et trois font cinq.

2. Où est l'église?

    1. J'y vais le dimanche.
    2. l'église est tout droit.
    3. Elles sont près d'ici.

3. Quel temps fait-il?

    1. Il est cinq heures.
    2. Il fait du français.
    3. Il fait mauvais.

4. Où va-t-il?

    1. Il va chez Philippe.
    2. Il y va tout de suite.
    3. J'y vais aussi.

5. Qu'est-ce que tu fais après l'école?

    1. Pas très bien.
    2. Ça ne fait rien.
    3. Je vais chez moi.

6. Où est-ce qu'elle habite?

    1. J'habite loin d'ici.
    2. Elle y va tout de suite.
    3. Elle habite près de l'église.

7. Combien de places y a-t-il à cette table?

   1. La table est près de la fenêtre.
   2. Il y a deux places à cette table.
   3. Je n'aime pas la glace.

8. Où sont les jeunes filles?

   1. Elles sont les amies de ma soeur.
   2. Elles font du français.
   3. Elles sont près de la porte.

9. Combien de garçons faut-il pour jouer aux cartes?

   1. Ils n'aiment pas jouer aux cartes.
   2. Il faut quatre garçons.
   3. Trois et un font quatre.

10. Qui est la jeune fille là-bas?

    1. C'est une amie de Robert.
    2. Oui, la jeune fille est là-bas.
    3. Elle y va tout de suite.

STANFORD UNIVERSITY
SCHOOL OF EDUCATION

____
Student

____
Teacher School

READING TEST

Read the following paragraph and supply the missing words from the list to the right of the paragraph. You will notice that there are more words than blanks, so that you will have a number of words left over. You may not use a word more than once. When you have used a word, draw a line through it as it appears in the right hand column. You will have ten minutes to complete this page.

Robert ______________ chez Philippe pour écouter la radio, mais il n'a pas de ____________ parce que la radio ne ______________ pas. Alors, les deux garçons écoutent des disques et après ils ______________ aux cartes. A trois heures, ils ______________ faim et ils vont au restaurant pour ______________. Robert prend un coca-cola parce qu'il a ____________.

sont
manger
froid
jouer
ont
marche
chance
allons
soif
vont
mange
disques
va
jouent

STANFORD UNIVERSITY
SCHOOL OF EDUCATION

____
Student

____
Teacher School

GRAMMAR TEST

Rewrite the following sentences, changing whatever is plural to singular. Check your paper to be sure that you made all the necessary changes. You will have ten minutes to complete this page.

1. Mes soeurs n'aiment pas le riz.

____

2. Les garçons ont du papier.

____

3. Les blondes donnent un coup de téléphone à Philippe.

____

4. Mes amis font du français.

____

5. Les jeunes filles vont à la porte.

____

STANFORD UNIVERSITY
SCHOOL OF EDUCATION

Student ______________________

Teacher ______________ School

WRITING TEST (Base and Supplemental)

Answer the following questions to the best of your ability. Make your answers as long as you like, but be sure to use complete sentences. You will have fifteen minutes to complete this page.

1. Où vas-tu?

2. Où est-ce que tu habites?

3. Qu'est-ce qu'il y a à manger aujourd'hui?

4. Qu'est-ce que tu fais après l'école?

5. Pourquoi n'écoutes-tu pas tes disques?

STANFORD UNIVERSITY
SCHOOL OF EDUCATION

SPEAKING TEST (Tape script)

There are twelve questions in this speaking test. Please answer all questions fully, using complete sentences. You will hear each question only once, so please listen carefully.

1. Comment t'appelles-tu?
2. Combien font deux et deux?
3. Combien font cinq et six?
4. Combien font trois et neuf?
5. Quel âge as-tu?
6. Quel temps fait-il?
7. Tu habites loin d'ici ou près d'ici?
8. Tu aimes mieux écouter des disques ou faire du ski?
9. Qu'est-ce que tu regardes après l'école, le journal ou la télévision?

In answering the last three questions, feel free to say whatever you like and as much as you like. You will be given extra time to make your responses.

10. Pourquoi vas-tu à la bibliothèque après l'école?
11. Qu'est-ce qu'il y a à manger aujourd'hui?
12. Qu'est-ce que tu fais le dimanche?

APPENDIX IX

Correlation Matrices

APPENDIX IX-a

Block AA
Intercorrelations of Teacher Behaviors

	(1) Prop. Dr. Time	(2) Ave. Dr./min.	(3) Rep. dr. pdm	(4) Sub. dr. pdm	(5) Dial. dr. pdm	(6)Trans. dr. pdm	(7) Conv. dr. pdm	(8) Free resp. pdm	(9) Ratio Cont/Free	(10)Ratio Sw./Excl.	(11) Ref. to book	(12) Use Vis. Aids	(13)St/St Interac.	(14)Var'n of Struc.
(1) Prop. Dr. Time	1.00 (17)													
(2) Ave. Dr./min.	.25 (17)	1.00 (17)												
(3) Rep. dr. pdm	-.12 (17)	$.45^{a}$ (17)	1.00 (17)											
(4) Sub. dr. pdm	.35 (17)	.51* (17)	-.28 (17)	1.00 (17)										
(5) Dial. dr. pdm	-.05 (17)	-.05 (17)	-.54* (17)	.14 (17)	1.00 (17)									
(6)Trans. dr. pdm	.14 (17)	.01 (17)	$-.44^{a}$ (17)	.19 (17)	.16 (17)	1.00 (17)								

(7) Conv. dr. pdm	.26 (17)	.41[a] (17)	.07 (17)	.10 (17)	-.11 (17)	-.33 (17)	1.00 (17)							
(8) Free resp. pdm	.26 (17)	.52* (17)	.24 (17)	.06 (17)	-.18 (17)	-.09 (17)	.44[a] (17)	1.00 (17)						
(9) Ratio Cont/Free	.21 (17)	.15 (17)	-.04 (17)	-.06 (17)	-.13 (17)	-.35 (17)	.95** (17)	.36 (17)	1.00 (17)					
(10)Ratio Sw./Excl.	.18 (17)	.35 (17)	.33 (17)	-.15 (17)	-.36 (17)	-.18 (17)	.52* (17)	.46[a] (17)	.47* (17)	1.00 (17)				
(11) Ref. to book	.20 (17)	-.06 (17)	.01 (17)	.16 (17)	.26 (17)	-.06 (17)	-.28 (17)	-.29 (17)	-.26 (17)	-.37 (17)	1.00 (17)			
(12) Use Vis. Aids	.21 (17)	.33 (17)	.18 (17)	-.03 (17)	-.15 (17)	-.28 (17)	.40[a] (17)	.66** (17)	.32 (17)	.46[a] (17)	-.39 (17)	1.00 (17)		
(13)St/St Interac.	.18 (17)	.05 (17)	-.38 (17)	.15 (17)	.55* (17)	-.05 (17)	.34 (17)	.08 (17)	.32 (17)	-.07 (17)	-.14 (17)	.31 (17)	1.00 (17)	
(14)Var'n of Struc.	.55* (17)	.42[a] (17)	-.03 (17)	.22 (17)	-.10 (17)	-.08 (17)	.57* (17)	.78** (17)	.50* (17)	.46[a] (17)	-.26 (17)	.62** (17)	.25 (17)	1.00 (17)

Notes to Appendix IX-a through IX-j:

Throughout these tables, the number in parentheses below the correlation coefficient indicates the size of the sample.

a Approaching significance
* $p < .05$
** $p < .01$

APPENDIX IX-b

Block BB
Intercorrelations of Teacher Characteristics

	(15) Attit. textbook	(16) Tchng Exper.	(17) Time in France	(18)MLA-ETS Listening	(19)MLA-ETS Speaking	(20)MLA-ETS Reading	(21)MLA-ETS Writing	(22)MLA-ETS Lingstcs	(23)MLA-ETS Cult.&Civ.	(24)MLA-ETS Prof. Prep.
(15) Attit. textbook	1.00 (17)									
(16) Tchng Exper.	.42[a] (17)	1.00 (17)								
(17) Time in France	.20 (17)	-.07 (17)	1.00 (17)							
(18)MLA-ETS Listening	-.27 (16)	-.26 (16)	.44[a] (16)	1.00 (16)						
(19)MLA-ETS Speaking	-.40 (15)	-.31 (15)	.21 (15)	.60* (15)	1.00 (15)					
(20)MLA-ETS Reading	-.56* (16)	-.14 (16)	.25 (16)	.67** (16)	.64** (15)	1.00 (16)				
(21)MLA-ETS Writing	-.51* (16)	-.31 (16)	.23 (16)	.69** (16)	.54* (15)	.83** (16)	1.00 (16)			
(22)MLA-ETS Lingstcs	-.56* (16)	-.36 (16)	.02 (16)	.45[a] (16)	.64** (15)	.66** (16)	.64** (16)	1.00 (16)		
(23)MLA-ETS Cult.&Civ.	-.43[a] (16)	.11 (16)	-.06 (16)	.48[a] (16)	.37 (15)	.71** (16)	.57* (16)	.79** (16)	1.00 (16)	
(24)MLA-ETS Prof. Prep.	-.50* (16)	-.03 (16)	-.31 (16)	.18 (16)	.07 (15)	.38 (16)	.39 (16)	.58* (16)	.63** (16)	1.00 (16)

APPENDIX IX-c

Block CC
Intercorrelations of Student Attitudes

	(25) Attit. pre-test	(26) Attit. post-test	(27) Gain [(26)-(25)]	(28)Int.to Cont. pre	(29)Int.to Cont. post	(30) Gain [(29)-(28)]	(31)Imp.of subj. pre	(32)Imp.of subj. post	(33) Gain [(32)-(31)]
(25) Attit. pre-test	1.00 (16)								
(26) Attit. post-test	.40[a] (16)	1.00 (16)							
(27) Gain [(26)-(25)]	-.53* (16)	.56* (16)	1.00 (16)						
(28)Int.to Cont. pre	.65** (16)	.72** (16)	.08 (16)	1.00 (16)					
(29)Int.to Cont. post	.42[a] (16)	.73** (16)	.30 (16)	.74** (16)	1.00 (16)				
(30) Gain [(29)-(28)]	-.40 (16)	-.11 (16)	.26 (16)	-.51* (16)	.20 (16)	1.00 (16)			
(31)Imp.of subj. pre	.71** (16)	.16 (16)	-.50* (16)	.18 (16)	-.10 (16)	-.40 (16)	1.00 (16)		
(32)Imp.of subj. post	.49[a] (16)	.39 (16)	-.08 (16)	.29 (16)	.15 (16)	-.25 (16)	.58* (16)	1.00 (16)	
(33) Gain [(32)-(31)]	-.42[a] (16)	.16 (16)	.53* (16)	.04 (16)	.26 (16)	.25 (16)	-.69** (16)	.19 (16)	1.00 (16)

APPENDIX IX-d

Block DD
Intercorrelations of Criterion Measures of Student Achievement

	(34)List'ng	(35)Reading	(36)Grammar	(37) Base Writing	(38) Free Writing	(39) Base Speaking	(40) Free Speaking
(34)List'ng	1.00 (17)						
(35)Reading	.53* (17)	1.00 (17)					
(36)Grammar	.49* (17)	.84** (17)	1.00 (17)				
(37) Base Writing	.66** (17)	.63** (17)	.68** (17)	1.00 (17)			
(38) Free Writing	.65** (17)	.47[a] (17)	.53* (17)	.86** (17)	1.00 (17)		
(39) Base Speaking	.58* (17)	.59* (17)	.79** (17)	.69** (17)	.77** (17)	1.00 (17)	
(40) Free Speaking	.52* (17)	.14 (17)	.32 (17)	.55* (17)	.74** (17)	.56* (17)	1.00 (17)

APPENDIX IX-e

Block AB
Correlations of Teacher Behaviors and Characteristics

	(15) Attit. textbook	(16) Tchng Exper.	(17) Time in France	(18)MLA-ETS Listening	(19)MLA-ETS Speaking	(20)MLA-ETS Reading	(21)MLA-ETS Writing	(22)MLA-ETS Lingstcs	(23)MLA-ETS Cult.&Civ.	(24)MLA-ETS Prof. Prep.
(1) Prop. Dr. Time	-.45[a] (17)	-.05 (17)	-.07 (17)	.03 (16)	.18 (15)	.25 (16)	.14 (16)	.46[a] (16)	.33 (16)	.29 (16)
(2) Ave. Dr./min.	-.20 (17)	-.22 (17)	.30 (17)	.57* (16)	.39 (15)	.05 (16)	.22 (16)	.22 (16)	.07 (16)	.17 (16)
(3) Rep. dr. pdm	-.11 (17)	-.29 (17)	.13 (17)	.21 (16)	.16 (15)	.08 (16)	.44[a] (16)	.36 (16)	.07 (16)	.40[a] (16)
(4) Sub. dr. pdm	.07 (17)	.18 (17)	-.09 (17)	.32 (16)	.07 (15)	-.12 (16)	-.15 (16)	.06 (16)	.27 (16)	.08 (16)
(5) Dial. dr. pdm	.38 (17)	.68** (17)	.02 (17)	-.25 (16)	-.05 (15)	-.26 (16)	-.56* (16)	-.50* (16)	-.23 (16)	-.34 (16)
(6)Trans. dr. pdm	-.26 (17)	-.20 (17)	-.23 (17)	.00 (16)	.12 (15)	-.04 (16)	-.29 (16)	-.14 (16)	-.15 (16)	-.07 (16)

(7) Conv. dr. pdm	-.20 (17)	-.37 (17)	.39 (17)	.12 (16)	.06 (15)	-.10 (16)	.00 (16)	.00 (16)	-.32 (16)	-.25 (16)
(8) Free resp. pdm	-.17 (17)	-.27 (17)	.73** (17)	.44[a] (16)	.23 (15)	.15 (16)	.30 (16)	.23 (16)	.02 (16)	-.15 (16)
(9) Ratio Cont/Free	-.22 (17)	-.34 (17)	.35 (17)	-.09 (16)	-.12 (15)	-.12 (16)	-.03 (16)	-.08 (16)	-.37 (16)	-.29 (16)
(10)Ratio Sw./Excl.	-.57[a] (17)	-.45[a] (17)	.16 (17)	.39 (16)	.24 (15)	.30 (16)	.40 (16)	.29 (16)	.14 (16)	.06 (16)
(11) Ref. to book	.16 (17)	.41[a] (17)	-.49* (17)	-.54* (16)	-.50[a] (15)	-.45[a] (16)	-.43[a] (16)	-.04 (16)	.12 (16)	.37 (16)
(12) Use Vis. Aids	-.13 (17)	-.24 (17)	.47[a] (17)	.58* (16)	.32 (15)	.36 (16)	.52[a] (16)	.25 (16)	.09 (16)	-.04 (16)
(13)St/St Interac.	.25 (17)	.45[a] (17)	.25 (17)	-.06 (16)	-.21 (15)	-.16 (16)	-.26 (16)	-.58* (16)	-.43[a] (16)	-.41[a] (16)
(14)Var'n of Struc.	-.18 (17)	-.28 (17)	.44[a] (17)	.25 (16)	.21 (15)	.07 (16)	.22 (16)	.16 (16)	-.06 (16)	-.32 (16)

APPENDIX IX-f

Block AC
Correlations of Teacher Behaviors and Student Attitudes

	(25) Attit. pre-test	(26) Attit. post-test	(27) Gain [(26)-(25)]	(28)Int. to Cont. pre	(29)Int. to Cont. post	(30) Gain [(29)-(28)]	(31)Imp.of subj. pre	(32)Imp.of subj. post	(33) Gain [(32)-(31)]
(1) Prop. Dr. Time	.37 (16)	.05 (16)	-.29 (16)	.30 (16)	.14 (16)	-.27 (16)	.40[a] (16)	.42[a] (16)	-.11 (16)
(2) Ave. Dr./min.	-.18 (16)	.14 (16)	.29 (16)	.21 (16)	-.08 (16)	-.40 (16)	-.05 (16)	-.08 (16)	.00 (16)
(3) Rep. dr. pdm	-.52* (16)	-.18 (16)	.31 (16)	-.32 (16)	-.24 (16)	.21 (16)	-.52* (16)	-.36 (16)	.31 (16)
(4) Sub. dr. pdm	.22 (16)	.10 (16)	-.11 (16)	.37 (16)	.27 (16)	-.21 (16)	.18 (16)	.04 (16)	-.18 (16)
(5) Dial. dr. pdm	.03 (16)	-.09 (16)	-.12 (16)	.01 (16)	-.15 (16)	-.27 (16)	.19 (16)	.33 (16)	.07 (16)
(6)Trans. dr. pdm	.48 (16)	.50* (16)	.03 (16)	.38 (16)	.26 (16)	-.27 (16)	.56* (16)	.52* (16)	-.22 (16)

(7) Conv. dr. pdm	-.09 (16)	-.26 (16)	-.16 (16)	-.19 (16)	-.48[a] (16)	-.30 (16)	.19 (16)	-.21 (16)	-.42[a] (16)
(8) Free resp. pdm	.12 (16)	.29 (16)	.16 (16)	.31 (16)	.17 (16)	-.22 (16)	-.01 (16)	-.10 (16)	-.07 (16)
(9) Ratio Cont/Free	-.02 (16)	-.30 (16)	-.26 (16)	-.24 (16)	-.48[a] (16)	-.25 (16)	.23 (16)	-.24 (16)	-.49* (16)
(10)Ratio Sw./Excl.	-.33 (16)	.07 (16)	.36 (16)	-.10 (16)	-.22 (16)	-.10 (16)	.02 (16)	-.12 (16)	-.13 (16)
(11) Ref. to book	.08 (16)	-.62** (16)	-.65** (16)	-.13 (16)	-.32 (16)	-.24 (16)	.04 (16)	.14 (16)	.07 (16)
(12) Use Vis. Aids	-.02 (16)	.31 (16)	.31 (16)	.24 (16)	.06 (16)	-.24 (16)	-.08 (16)	.14 (16)	.22 (16)
(13)St/St Interac.	.26 (16)	.03 (16)	-.20 (16)	-.03 (16)	-.14 (16)	-.12 (16)	.46[a] (16)	.41[a] (16)	-.20 (16)
(14)Var'n of Struc.	.38 (16)	.27 (16)	-.10 (16)	.50[a] (16)	.17 (16)	-.48[a] (16)	.27 (16)	.00 (16)	-.32 (16)

APPENDIX IX-g

Block BC
Correlations of Teacher Characteristics and Student Attitudes

	(15) Attit. textbook	(16) Tchng Exper.	(17) Time in France	(18)MLA-ETS Listening	(19)MLA-ETS Speaking	(20)MLA-ETS Reading	(21)MLA-ETS Writing	(22)MLA-ETS Lingstcs	(23)MLA-ETS Cult.&Civ.	(24)MLA-ETS Prof. Prep.
(25) Attit. pre-test	.10 (16)	-.08 (16)	-.15 (16)	-.35 (15)	-.22 (14)	-.42^{a} (15)	-.34 (15)	-.28 (15)	-.42^{a} (15)	-.40 (15)
(26) Attit. post-test	-.11 (16)	-.29 (16)	.24 (16)	.56* (15)	.57* (14)	.35 (15)	.33 (15)	.09 (15)	-.02 (15)	-.27 (15)
(27) Gain [(26)-(25)]	-.19 (16)	-.20 (16)	.35 (16)	.82** (15)	.67* (14)	.70** (15)	.61* (15)	.33 (15)	.36 (15)	.12 (15)
(28)Int.to Cont. pre	-.05 (16)	-.18 (16)	.01 (16)	.28 (15)	.40 (14)	.05 (15)	.15 (15)	.14 (15)	.03 (15)	-.34 (15)
(29)Int.to Cont. post	.04 (16)	.04 (16)	.13 (16)	.38 (15)	.43^{a} (14)	.35 (15)	.32 (15)	.30 (15)	.33 (15)	-.19 (15)
(30) Gain [(29)-(28)]	.14 (16)	.29 (16)	.17 (16)	.07 (15)	-.09 (14)	.35 (15)	.21 (15)	.14 (15)	.33 (15)	.22 (15)
(31)Imp.of subj. pre	-.17 (16)	-.07 (16)	-.30 (16)	-.41^{a} (15)	-.30 (14)	-.45^{a} (15)	-.51^{a} (15)	-.44^{a} (15)	-.52* (15)	-.22 (15)
(32)Imp.of subj. post	-.02 (16)	.08 (16)	-.22 (16)	-.11 (15)	.08 (14)	-.08 (15)	-.30 (15)	-.11 (15)	-.18 (15)	.01 (15)
(33) Gain [(32)-(31)]	.19 (16)	.15 (16)	.17 (16)	.45^{a} (15)	.46^{a} (14)	.54* (15)	.40^{a} (15)	.48^{a} (15)	.53* (15)	.31 (15)

APPENDIX IX-h

Block AD
Correlations of Teacher Behaviors and Criterion Measures of Student Achievement

	(34) List'ng	(35) Reading	(36) Grammar	(37) Base Writing	(38) Free Writing	(39) Base Speaking	(40) Free Speaking
(1) Prop. Dr. Time	-.34 (17)	-.37 (17)	-.33 (17)	-.15 (17)	-.07 (17)	-.26 (17)	-.01 (17)
(2) Ave. Dr./min.	-.01 (17)	.14 (17)	.04 (17)	.35 (17)	.45[a] (17)	.33 (17)	.20 (17)
(3) Rep. dr. pdm	-.08 (17)	.31 (17)	.21 (17)	.24 (17)	.22 (17)	.28 (17)	.01 (17)
(4) Sub. dr. pdm	.08 (17)	-.04 (17)	-.22 (17)	.13 (17)	.17 (17)	-.04 (17)	-.06 (17)
(5) Dial. dr. pdm	-.21 (17)	-.39 (17)	-.43[a] (17)	-.58* (17)	-.35 (17)	-.26 (17)	-.18 (17)
(6) Trans. dr. pdm	-.18 (17)	-.22 (17)	-.31 (17)	-.10 (17)	-.14 (17)	-.49* (17)	.14 (17)
(7) Conv. dr. pdm	.28 (17)	.05 (17)	.30 (17)	.42[a] (17)	.40[a] (17)	.45[a] (17)	.19 (17)
(8) Free resp. pdm	.15 (17)	.06 (17)	.20 (17)	.61** (17)	.73** (17)	.50* (17)	.62** (17)
(9) Ratio Cont/Free	.26 (17)	.02 (17)	.33 (17)	.37 (17)	.35 (17)	.40[a] (17)	.16 (17)
(10) Ratio Sw./Excl.	.28 (17)	.25 (17)	.53* (17)	.46[a] (17)	.55* (17)	.65** (17)	.68** (17)
(11) Ref. to book	-.46[a] (17)	-.20 (17)	-.29 (17)	-.42[a] (17)	-.34 (17)	-.30 (17)	-.61** (17)
(12) Use Vis. Aids	.26 (17)	-.08 (17)	.15 (17)	.36 (17)	.48[a] (17)	.57* (17)	.59* (17)
(13) St/St Interac.	.14 (17)	-.50* (17)	-.47[a] (17)	-.28 (17)	-.09 (17)	-.09 (17)	.07 (17)
(14) Var'n of Struc.	.03 (17)	.00 (17)	.17 (17)	.42[a] (17)	.42[a] (17)	.35 (17)	.39 (17)

APPENDIX IX-i

Block BD
Correlations of Teacher Characteristics and Criterion Measures of Student Achievement

	(34)List'ng	(35)Reading	(36)Grammar	(37) Base Writing	(38) Free Writing	(39) Base Speaking	(40) Free Speaking
(15) Attit. textbook	.21 (17)	.04 (17)	-.18 (17)	-.07 (17)	-.19 (17)	-.12 (17)	-.35 (17)
(16) Tchng Exper.	-.11 (17)	-.32 (17)	-.49* (17)	-.59* (17)	-.29 (17)	-.25 (17)	-.26 (17)
(17) Time in France	.43[a] (17)	.18 (17)	.18 (17)	.52* (17)	.61** (17)	.40[a] (17)	.47[a] (17)
(18)MLA-ETS Listening	.41[a] (16)	.31 (16)	.29 (16)	.49[a] (16)	.60* (16)	.54* (16)	.63** (16)
(19)MLA-ETS Speaking	-.24 (15)	-.06 (15)	.02 (15)	-.01 (15)	.00 (15)	.04 (15)	.29 (15)
(20)MLA-ETS Reading	.15 (16)	.18 (16)	.22 (16)	.08 (16)	.20 (16)	.22 (16)	.43[a] (16)
(21)MLA-ETS Writing	.11 (16)	.34 (16)	.39 (16)	.29 (16)	.34 (16)	.44[a] (16)	.37 (16)
(22)MLA-ETS Lingstcs	-.17 (16)	.11 (16)	.18 (16)	.16 (16)	.15 (16)	.12 (16)	.15 (16)
(23)MLA-ETS Cult.&Civ.	.03 (16)	.18 (16)	.16 (16)	.04 (16)	.19 (16)	.18 (16)	.17 (16)
(24)MLA-ETS Prof. Prep.	-.13 (16)	.07 (16)	-.06 (16)	-.11 (16)	.08 (16)	-.03 (16)	-.08 (16)

APPENDIX IX-j

Block CD
Correlations of Student Attitudes and Criterion Measures of Student Achievement

	(34)List'ng	(35)Reading	(36)Grammar	(37) Base Writing	(38) Free Writing	(39) Base Speaking	(40) Free Speaking
(25) Attit. pre-test	-.38 (16)	-.50* (16)	-.50* (16)	-.10 (16)	-.25 (16)	-.54* (16)	-.15 (16)
(26) Attit. post-test	.00 (16)	-.12 (16)	-.14 (16)	.16 (16)	.13 (16)	-.10 (16)	.52* (16)
(27) Gain [(26)-(25)]	.34 (16)	.34 (16)	.32 (16)	.24 (16)	.34 (16)	.40 (16)	.62* (16)
(28)Int.to Cont. pre	-.42[a] (16)	-.24 (16)	-.17 (16)	.05 (16)	-.06 (16)	-.14 (16)	.10 (16)
(29)Int.to Cont. post	-.14 (16)	-.16 (16)	-.24 (16)	.04 (16)	-.03 (16)	-.26 (16)	.20 (16)
(30) Gain [(29)-(28)]	.48[a] (16)	.19 (16)	-.01 (16)	.03 (16)	.08 (16)	-.06 (16)	.12 (16)
(31)Imp.of subj. pre	-.21 (16)	-.51* (16)	-.48[a] (16)	-.20 (16)	-.18 (16)	-.45[a] (16)	.04 (16)
(32)Imp.of subj. post	-.36 (16)	-.74** (16)	-.76** (16)	-.50* (16)	-.40 (16)	-.58* (16)	.06 (16)
(33) Gain [(32)-(31)]	-.07 (16)	-.04 (16)	-.10 (16)	-.20 (16)	-.14 (16)	.02 (16)	.01 (16)